PHYSICAL MAPPING AND IDENTIFICATION OF GENES FROM A REGION OF CHROMOSOME 11 ADJACENT TO A TRANSLOCATION BREAKPOINT ASSOCIATED WITH SCHIZOPHRENIA

Rebecca S. Devon

PhD

The University of Edinburgh

1995

DECLARATION

I declare that:

a) this thesis has been composed by myself

b) that the work is my own, except where otherwise stated

Rebecca S. Devon

October 1995

ACKNOWLEDGEMENTS

First of all I would like to thank wholeheartedly my supervisors David Porteous and Tony Brookes. I have been extremely fortunate during my PhD to have had such encouraging and supportive supervisors, and the combination of their, sometimes differing, advice usually provided me with the right answer. I would like to thank David particularly for his enthusiasm and foresight, for pretending to smile as I handed him another chapter to read and for the Friday afternoon lab drinks parties. I would like to thank Tony particularly for being so patient when I pestered him constantly (and in my first year I mean *constantly*), for teaching me that clean Gilsons really *are* important, for giving me faith in my own work and for his annoying habit of always being right.

Thanks also to everyone in the West Wing for making it such a great place to work - so full of life, laughter and innnovation (both scientific and otherwise). That goes especially to the girls in the "Super-bay" (DJP), to Kathy for moral support and lots of fun, to Euan for surreal (no, ridiculous) conversations and teaching me how to eject pipette tips with style, and to Sheila whose efficiency and stoicism in the face of broken equipment and continual demands for enzymes never fails to amaze me. Thanks also to Jane for going through everything at the same time as me and for those long Sunday brunches.

Many others also deserve my thanks, particularly Sheila Mould for uncomplainingly ordering hundreds of papers for me, and Sandy Bruce for his cheerful preparation of all the figures here. Elena Farini, Kathy Williamson and Andreas Schedl kindly provided me with some primers and probes.

My final thanks of course has to go to my family and friends, without whom I'd have probably become psychotic myself. Thanks to Giorgia and Steve for being such excellent flatmates and also to Mum, Dad, James and especially to Donald for their support, care and for always being there for me.

ABSTRACT

Schizophrenia is the most severe of all the psychoses, affecting approximately 1% of the world's population. Although the aetiology of the disorder is likely to be multifactorial, evidence accumulated from numerous family, twin and adoption studies has firmly established a significant genetic component. The identification of the relevant genes by classical linkage analysis has been plagued by problems such as genetic heterogeneity and polygenic inheritance, and as such no decisive and replicable gene/disorder associations have yet been detected. The densely spaced markers now available throughout much of the human genome have facilitated whole genome positional cloning efforts. However the presence of a cytogenetic rearrangement associated with the disorder provides an extremely valuable research tool. A balanced t(1;11) translocation has been described which co-segregates with a broad range of related psychiatric conditions, including schizophrenia, in a large Scottish family. Generation and mapping of markers in the region of this translocation led to the construction of a 3Mb yeast artificial chromosome (YAC) contig across the chromosome 11 breakpoint. This thesis begins with the isolation of YAC end clones by 'splinkerette PCR', an improved alternative to vectorette PCR, which facilitated the contig construction. A search for foetal brain-expressed genes from this YAC was then undertaken, using two complementary methods based on cDNA selection, ('hybrid fishing' and 'end ligation' coincident sequence cloning). A thorough analysis of the product library and confirmatory hybridisation to the YAC and cDNA led to the identification of an additional member of the α-tubulin gene family, and several novel gene fragments, comprising at least two genes. Extensive sequence and expression analysis of the α-tubulin gene suggest that it is a processed pseudogene, although its potential as a candidate gene for psychiatric illness in this family cannot be discounted. The novel gene fragments have been mapped relative to each other and to rare cutter restriction sites in cloned genomic DNA, allowing one group of four fragments to be tentatively assigned to a single CpG island-associated gene. These can now be tested as genetic susceptibility factors in schizophrenia.

ABBREVIATIONS USED

µCi	microcuries
µM	micromolar
µl	microlitres
5HT	serotonin
A	adenine
abs	absorbance
ACh	acetyl choline
AP-SA	alkaline phosphatase-streptavidin
APP	amyloid precursor protein
BAC	bacterial artificial chromosome
BCIP	5-bromo-4-chloro-3-indolylphosphate
bp	base pair
BSA	bovine serum albumin
C	cytosine
cAMP	cyclic adenosine monophosphate
$CaPO_4$	calciom phosphate
CCK	cholecystokinin
cDNA	complementary deoxyribonucleic acid
CEPH	Centre d'Étude du Polymorphisme Humain
CFTR	cystic fibrosis transmembrane regulator
cfu	colony forming unit
CIP	calf intestinal phosphatase
cM	centimorgan
CSC	coincident sequence cloning
CsCl	caesium chloride
CSF	cerebrospinal fluid
CT	computed tomography
D1-5	dopamine receptor 1-5

dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
DEPC	diethyl pyrocarbonate
DGGE	denaturing gradient gel electrophoresis
dGTP	deoxyguanosine triphosphate
dH_20	distilled water
DMF	dimethylformamide
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DSM	American Diagnostic and Statistical Manual
DTT	dithiothreitol
dTTP	deoxythymine triphosphate
DZ	dizygotic
E.Coli	Escherischia Coli
ECT	electroconvulsive therapy
EDTA	ethylenediamine tetra-acetic acid disodium salt
EE	expressed emotion
EFA	essential fatty acid
EL	end ligation
ERP	event related potential
EST	expressed sequence tag
EtBr	ethidium bromide
FAS	foetal alcohol syndrome
FISH	fluorescence *in-situ* hybridisation
FRS	First Rank Symptoms
G	guanosine
G6PD	glucose-6-phosphate dehydrogenase
GABA	gamma amino butyric acid
GAPDH	glutaraldehyde phosphate dehydrogenase
GCG	genetics computer group
GFAP	glial fibrillary acidic protein
HCl	hydrochloric acid

HD	Huntington's disease
HF	hybrid fishing
HGMP	human genome mapping project
HLA	human lymphocyte antigen
hnRNA	heteronuclear ribonucleic acid
HOT	hydroxylamine osmium tetroxide
HTF	HpaII tiny fragments
HVA	homovanillic acid
IBD	identity by descent
ICD	International Classification of Diseases
IDDM	insulin-dependent diabetes mellitus
IPTG	isopropylthio-β-D-galactoside
IQ	intelligence quotient
IRD	inter resource duplex
K_2HPO_4	dipotassium hydrogen phosphate
KAc	potassium acetate
kb	kilobase
KH_2PO_4	potassium dihydrogen phosphate
kV	kilovolts
l	litre
LMP	low melting point
LOD	logarithm of odds
LSD	lysergic acid diethylamide
m	milli
M	molar
mA	milliamps
MAO	monoamine oxidase
MAP	microtubule-associated protein
Mb	megabase
MEG	magnetoencephalography
MFP	multifactorial polygenic
$MgCl_2$	magnesium chloride

$MgSO_4$	magnesium sulphate
MPA	minor physical abnormality
MRC	Medical Research Council
MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid
MRS	magnetic resonance spectroscopy
MZ	monozygotic
n	nano
Na_3Ci	sodium citrate
NaAc	sodium acetate
NaCl	sodium chloride
NaOH	sodium hydroxide
NBT	nitroblue tetrazolium
NCAM	neural cell adhesion molecule
NH_4SO_4	ammonium sulphate
NMDA	N-methyl-D-aspartate
NOD	non-obese diabetic
OCA	oculocutaneous albinism
ORF	open reading frame
p	pico
PAC	P1 artificial chromosome
PBGD	porphobilonogen deaminase
pBS	Bluescribe plasmid vector
PCP	phencyclidine
PCR	polymerase chain reaction
PEA	phenylethylamine
PEG	polyethylene glycol
PET	positron emission tomography
PFGE	pulsed field gel electrophoresis
PIC	polymorphism information content
PKU	phenylketonuria
PMSF	phenylmethylsulphonyl fluoride

PNK	polynucleotide kinase
Prep-ISH	Preparative In Situ Hybridisation
PVP	polyvinylpyrolidone
QTL	quantitative trait locus/loci
RDC	Research Diagnostic Criteria
RED	repeat expansion detection
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
RT	room temperature
RT-PCR	reverse transcriptase polymerase chain reaction
SDS	sodium dodecyl sulphate
SML	single major locus
SSCP	single strand conformation polymorphism
SSR	simple sequence repeat
STS	sequence tagged site
T	thymine
TCA	trichloroacetic acid
TE	10mM Tris.HCl (pH 8 unless otherwise stated), 1mM EDTA
TEMED	N,N,N,N'-tetramethylethylenediamine
TH	tyrosine hybdroxylase
T_m	melting temperature
TYR	tyrosinase
UTR	untranslated region
V	volts
VNTR	variable number of tandem repeats
WT1	Wilm's Tumour 1 gene
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
YAC	yeast artificial chromosome
$ZnCl_2$	zinc chloride

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

1.1 Preface

Schizophrenia is one of the most severe of all mental illnesses. It is not a 'split mind' as is commonly believed, but rather a disorder characterised by mental anguish or insanity, sometimes so debilitating that the sufferer permanently loses all touch with reality. It affects one in every hundred people world-wide, and its sufferers fill nearly a quarter of the hospital beds in Britain (Wallace, 1987). One third of the prison population in the UK suffers from the disorder. Mental illness accounts for 14% of certified sickness absence, 14% of the National Health Service (NHS) Inpatient costs and 23% of the NHS pharmaceutical costs (Medical Research Council Field Review on Biological Psychiatry, 1993). More than one third of the urban homeless in the UK have had contact with psychiatric services.

Despite many decades of research into schizophrenia, little is established concerning the cause. Modern treatments do little more than ameliorate the most bizarre symptoms, leaving many schizophrenics unable to cope in the community and in a downward spiral of social isolation. There is convincing evidence however that schizophrenia is, at least partly, a genetic disease. So, as the pace of human genome research gathers speed, and knowledge of molecular mechanisms increases, there is now a real possibility of gaining an understanding of the underlying causes of the disorder.

1.2 Diagnosis of Schizophrenia

Although schizophrenia is probably an innate part of human existence, it was not described until the beginning of the nineteenth century, by Pinel in 1798 and Haslam in 1809 (Wyatt, 1988). Kraepelin is credited with the first

diagnostic description of the disease, 'dementia praecox' as he termed it, as having a chronic course with progressive deterioration. Somewhat later, Bleuler coined the modern day terminology when he described 'the group of schizophrenias' (Wyatt, 1988).

Since these first descriptions of schizophrenia, there has been considerable debate over how it should be correctly diagnosed. Schizophrenia is a heterogeneous disorder both in clinical presentation and outcome, and as such is sometimes difficult to classify. There have been many attempts to devise reliable diagnostic systems, such as the American Diagnostic and Statistical Manual (DSM), the Research Diagnostic Criteria (RDC) and the International Classification of Diseases (ICD). Standardisation of diagnosis by these structured interview schedules is most certainly beneficial. Diagnoses of schizophrenia usually concur between the different diagnostic systems, although they do differ in their treatment of fringe phenotypes.

1.3 Epidemiology of Schizophrenia

Schizophrenia has a rate in the general population of approximately 1% (Wyatt, 1988). What is remarkable about a disorder so heterogeneous is that this prevalence rate is virtually constant in all geographical regions it has been tested, although there are a few isolated pockets of high incidence, for instance Northern Sweden, Ireland and Finland (Wyatt, 1988). The fact that the rate remains constant across widely varying cultures in developed and developing nations is strongly suggestive that the underlying causes of schizophrenia are fundamental and intrinsic to man, and the disorder is not merely a result of cultural factors.

Of fifteen studies reviewed, ten have suggested that schizophrenia is a disorder of falling incidence (Harrison and Mason, 1993). On closer inspection however, it appears that these findings may be artefactual. The

apparent drop in incidence can probably be attributed to earlier intervention by general practitioners and therefore more prompt treatment, fewer false positives as a result of rigorous diagnostic practice and changes in the accepted diagnostic criteria themselves, with the inclusion of only those who have been ill for some time (Harrison and Mason, 1993).

1.4 Pathology of Schizophrenia

1.4.1 Symptoms

The most prevalent symptom in schizophrenia is lack of insight, which has been found to occur more frequently in schizophrenia than in the other psychoses (Cuesta and Peralta, 1994). The other symptoms of schizophrenia fall loosely into positive and negative categories. Positive symptoms include delusions, visual and auditory hallucinations (voices inside the head arguing, keeping up a running commentary or broadcasting one's thoughts), incoherent speech, positive thought disorder and inappropriate affect (Michels and Marzuk, 1993). Negative symptoms are characterised by social isolation, withdrawal, blunted affect, attentional impairment, loosening of associations, poverty of speech and lack of initiative, interests and energy (Rao and Moller, 1994).

To Bleuler, the four primary symptoms of the deficit state of schizophrenia were thought association problems, ambivalence, affective abnormalities and autistic thinking (Gallant *et al*, 1990). The more prominent symptoms of delusions and hallucinations were secondary symptoms. This still seems largely appropriate today, since the Schneiderian First Rank symptoms (FRS) such as delusions and hallucinations, upon which early diagnostic systems were based, have proved not to be pathognomic for schizophrenia (Gallant *et al*, 1990). Andreasen and Olsen (1982) found that the frequencies of purely positive symptoms, purely negative symptoms or a mixture of both were approximately equal - each constituting a third of cases. However long

term follow-up studies (Marneros *et al*, 1992) have found that few schizophrenics (18%) will have more than one episode of purely positive symptoms, and only 6% would experience exclusively negative symptoms. The vast majority would experience both during the course of their illness.

In the revised third version of the DSM (DSM IIIR), consideration is also given to the length of illness - the disorder must have been present for at least six months without a return to the premorbid level of psychosocial adjustment for schizophrenia to be diagnosed (Gallant *et al*, 1990). A brief episode of illness is diagnosed as schizophreniform disorder. The DSM IIIR also stipulates that there must be an absence of manic or depressive symptoms sufficient to qualify the individual for affective disorder (Gallant *et al*, 1990). The distinction between schizophrenia and affective disorder is highly contentious and will be discussed in section 1.4.3)

1.4.2 Subdivision of 'the group of schizophrenias'

The DSM IIIR recognises five principal subtypes of schizophrenia (Gallant *et al*, 1990).

The catatonic subtype is characterised by marked psychomotor disturbance such as stupor, negativism, rigidity, purposeless movement and posturing. The disorganised (or hebephrenic) subtype is recognised by incoherence, flat or inappropriate affect, loosening of associations, grimaces and mannerisms. There may be delusions and hallucinations but they are usually fragmentary and disorganised. The paranoid subtype includes at least one delusion and frequent auditory hallucinations related to one theme. Associated with these are excessive anxiety, anger, argumentativeness and sometimes violence. The undifferentiated subtype encompasses those disorders which appear to fit none or several of the other subtypes. Finally, a diagnosis of residual schizophrenia is given when there has been at least one active episode, but there is now no prominent psychosis. Residual

symptoms include emotional blunting, social withdrawal and eccentric behaviour.

Several studies have attempted to ascertain whether these subtypes are genetically distinct. Subtype is almost always consistent between two affected MZ twins (McGuffin *et al*, 1987), which does suggest subtype specificity in the inheritance of schizophrenia, if the finding is not due to ascertainment biases.

An alternative subdivision of schizophrenia (Crow, 1980a) is the Type I/ Type II split. Type I schizophrenia is characterised by mostly positive symptoms, good response to neuroleptic drugs, normal CT scans and a hyperdopaminergic aetiology. Type II schizophrenia however has mainly negative symptoms, is associated with brain atrophy and cognitive impairment and the aetiology is not dopaminergic.

A more radical proposal is to split schizophrenia into 'familial' and 'sporadic' subtypes, that is genetic versus non-genetic (Murray *et al*, 1985). The belief is that the genetic form is caused by an aberrant gene involved in brain development, and the sporadic form is caused by obstetric complications (in males) or auto-immunity (in females).

1.4.3 The schizophrenia spectrum

Psychiatric diagnoses other than schizophrenia are found to aggregate in the relatives of schizophrenics, and as a result many studies have attempted to define the extent of the schizophrenia 'spectrum'. The results are conflicting, which can largely be attributed to methodological discrepancies. A general consensus is that rates of schizoaffective disorder, schizotypal and paranoid personality disorders and atypical psychosis are all increased in the relatives of schizophrenics (Kendler *et al*, 1985). As such they can be tentatively ascribed to the same genetic mechanism and included in the spectrum.

Anxiety disorder, unipolar depression and alcoholism do not have a greater incidence in the families of schizophrenics and therefore are not included (Kendler *et al*, 1985).

The link with Bipolar Affective Disorder (Manic Depression)

There is an ongoing debate as to whether schizophrenia and affective disorder are entirely distinct disorders or instead represent the two extremes of a continuum of psychotic illness. Kraepelin believed that they were two separate disorders; schizophrenia implying progressive deterioration and affective disorder following a remitting course. The most recent, methodologically superior family studies, in support, conclude that rates of bipolar affective disorder are not increased in the relatives of schizophrenics. Similarly there is no excess of schizophrenia in the relatives of affectively ill probands (Baron and Gruen, 1991). However published case studies have described monozygotic twins, one with schizophrenia and one with bipolar disorder (Lohr and Bracha, 1992), and also a case of identical triplets where two were diagnosed with schizophrenia and one bipolar disorder (McGuffin *et al*, 1982). This could of course be a result of the unfortunate genetic presence of both illnesses in the family, but it seems more likely that the two diagnoses are a result of different manifestations of the same genetic defect.

Schizoaffective disorder, the co-existence of the symptoms of both schizophrenia and affective symptoms in one individual, is diagnosed as a discrete entity, but this implies nothing about its aetiology. Just as the other schizophrenia spectrum disorders aggregate in relatives of both schizophrenics and individuals with bipolar disorder, schizoaffective disorder may too be an alternative manifestation of either disorder. Alternatively, schizoaffective disorder may have an entirely different aetiology. Finally, the presence of both sets of symptoms in schizoaffective disorder, and its outcome intermediate between the two, has led many to believe that

schizoaffective disorder is a median diagnosis on a continuum from schizophrenia to affective disorder (Lapensée, 1992).

1.4.4 Age of onset

Schizophrenia is a disorder of late adolescence and adulthood. The usual age of onset is 18-25 years for men and 26-45 years for women (Wyatt, 1988). The disorganised subtype has the earliest onset, and paranoid schizophrenia the latest, with an average gap of thirteen years between them (Beratis *et al*, 1994). Between siblings, the *age* of onset is correlated rather than the *time* of onset, which is support for a genetic rather than environmental aetiology (deLisi, 1992).

1.4.5 Course and outcome

After onset, the course of schizophrenia is highly variable. Some patients will suffer one episode of psychosis and then return to near normal function, whereas others will experience an unremitting deterioration. In general however, schizophrenia is an episodic disease, with an average of more than five admissions to a psychiatric hospital (Johnstone, 1993). Unlike dementias, there is no evidence for progressive cognitive decline other than that linked to normal ageing (Hyde *et al*, 1994). A World Health Organisation (WHO) comparison of the course of schizophrenia across the world reveals that 10-28% will remain chronically psychotic, 10-64% will be free of symptoms at the time of follow-up, while 5-71% will exhibit 'good' social adjustment (Eaton, 1991). The wide variation in the last two figures is partly due to differing interpretation of terminology, but it is known that there is a better average prognosis for schizophrenics in developing countries. This is probably due to a requirement for fewer coping skills in order to function well in the community. Ten percent of schizophrenic patients commit suicide, although two thirds will attempt it.

The outcome figures quoted above are improved considerably since Kraepelin described dementia praecox, and there exists a general impression that the disorder is becoming more benign (Harrison and Mason, 1993). In the early nineteenth century, 19.5% of schizophrenics were classified as catatonic subtype, whereas in the 1970's it was only 5-10%. In Kraepelin's time, 59% of catatonics and 75% of hebephrenics reached a terminal state of profound dementia, with only 4.1% making a complete recovery (Wyatt, 1988). It is unclear whether the greatly improved prognosis today reflects a genuine progression to a milder form or is merely a result of changed diagnostic systems and improved treatment regimes. A substantial portion must be accredited to improved medications, but improvements in treatment are also due to greater emphasis on psychosocial awareness, such as teaching skills for symptom management, living and coping in the community. There is a greater emphasis too on family involvement - with education programmes, support and practical help (Torrey and Drake, 1994).

Estimates of good outcome are however confounded by the criteria used to define them and changes in diagnostic practice. Care must be exercised when interpreting outcome data across time. For instance the percentage of schizophrenics with a good outcome now is only the same as it was in the 1950's, despite modern therapeutic techniques (Hegarty, 1994). This is largely due to the adoption of the DSM IIIR for diagnosis, in which the inclusion criteria for schizophrenia specify a long course of illness.

1.4.6 Relapse and Expressed Emotion

Relapse rates reported for schizophrenia vary considerably, from 33% up to 80% after two years (Davies, 1994). In addition to the effectiveness of the maintenance medication, relapse rates are heavily dependent on the family circumstances. Family attitudes to the ex-patient are measured in terms of 'expressed emotion' (EE), which is assessed by a semi-structured interview of the family, rating factors such as hostility, warmth and emotional

overinvolvement (Bebbington and Kuipers, 1994). An high EE family member will have a tendency to generate or escalate conflict, and will be less effective at coping. They are generally of the view that the disorder is within the patient's control. A person of low EE will tend to avoid hostility and believes that there is nothing the patient can do to control his symptoms.

There is a conclusive body of evidence to suggest that a patient in contact with at least one family member of high EE (especially if that family member is particularly significant, such as a parent or spouse) has a vastly increased chance of relapse. Reciprocally, contact with a low EE family member can even have a protective effect against relapse (Bebbington and Kuipers, 1994).

1.4.7 Sex Differences

Although the prevalence of schizophrenia is approximately equal for both men and women, considerable variation is seen in the onset and course of the disorder. Males exhibit an earlier and sharper peak for age of onset (18-25 years) compared with females (26-45 years) (Wyatt, 1988). This has been proven not to be a result of excessive 'help seeking' behaviour in men or differences in social rôle expectations between the sexes (Castle and Murray, 1991). Males usually have more typical symptomatology, with more negative features and a tendency to withdrawal and passivity. They are also more likely to show poor pre-morbid social functioning and have lower pre-morbid IQ scores. In females the symptoms are usually characterised by a stronger affective component, and the course of the disorder is usually less severe than in males. Females respond better to neuroleptic medication, and have fewer and shorter stays in a psychiatric hospital (Castle and Murray, 1991).

Explanations proposed for these sex differences are increased rôle stress in males, protective rôles in females or hormonal differences (a triggering effect of androgens and/or a protective affect of oestrogens). Epidemiological studies have shown a second major peak of incidence for women, between ages 40 and 45, which may be due to the pre-menopausal drop in oestrogen levels. Also some studies have demonstrated changes in the psychopathology of schizophrenics at different stages of the menstrual cycle, with an exacerbation of symptoms seen when oestrogen levels are low (Riecher-Rössler *et al*, 1994). Administration of oestrogens can ameliorate psychosis, and in animals, chronic oestrogen treatment reduces the sensitivity of the D2 receptor in the brain (Riecher-Rössler *et al*, 1994).

There is a slower pace of cerebral development in males, which makes them statistically more prone to neurodevelopmental disorders in general. Obstetric complications such as periventricular haemorrhage and asphyxia are more likely during the birth of males (Castle and Murray, 1991). Obstetric complications are known to cause increased ventricular size and are correlated with an earlier age of onset of schizophrenia.

1.4.8 Reproductive fitness

It is generally accepted that schizophrenics have a reduced reproductive fitness, of approximately 70% compared with unaffected individuals (Wyatt, 1988). This may be a biological phenomenon or just a secondary effect, since the symptoms of the disorder may lessen the chances of marriage and cause breakdown in existing family relationships. How then have the deleterious mutations not been diluted out by natural selection? Assuming that the new mutation rate of schizophrenogenic genes is not sufficient to ensure their maintenance, some mechanism for their selective retention is implied.

From a genetic point of view, the simplest explanation is that many genes are involved in the disease process. If this were due to aetiological heterogeneity, then the normal new mutation rate may account for the observed frequency of schizophrenia, by acting at many different loci yet causing the same phenotype (Carter and Watts, 1971). If, on the other hand, the involvement of many genes was due to a polygenic mode of inheritance, the decrease in population frequency of an individual mutant allele would be slower than if the disorder were caused by a single gene, since a combination of mutant alleles is required before the deleterious phenotype is present. The gene(s) may not be fully penetrant, and there may be no loss of reproductive fitness in those with late onset (Ridley and Baker, 1990). Historically, most reproduction would have taken place before even the average age of onset of the phenotype, thus decreased reproductive fitness may be a recent phenomenon, the effects of which are not yet apparent.

In addition, schizophrenogenic gene mutations may confer a biological or sociological advantage on those carrying them. It is possible historically that those experiencing delusions and hallucinations were revered as 'visionaries' or 'mystics' and as such had a favourable place in society (Erlenmeyer-Kimling and Paradowski, 1966). This would hardly seem applicable today, yet the effect may not yet have been 'diluted out' by natural selection. The advantage may alternatively be conferred onto unaffected mutation carriers, the relatives of schizophrenics. This 'heterozygote advantage' may be in the form of resistance to infectious disease (for instance influenza, see section 1.6.5), such as is seen for heterozygosity for sickle cell anaemia conferring some resistance to malaria. One study found schizophrenics' relatives to have an incidence of viral infections of only 40% that expected, but no decrease in bacterial infections (Carter and Watts, 1971). It has also been proposed that the relatives of schizophrenics are unusually creative or intelligent people, however it is unclear whether this would confer a reproductive advantage on them. Erlenmeyer-Kimling (1968) proposed that

relatives of schizophrenics possess a genetically determined protection against the stresses and strains of social interaction. Another study (Erlenmeyer-Kimling and Paradowski, 1966) found an increase in reproductivity in the sisters of schizophrenics - 140% that of the normal population.

1.4.9 Premorbid functioning

The best predictor of prognosis after onset of schizophrenia is the level of premorbid functioning, as indicated by social ability and factors such as marital and occupational status. Poor social competence is correlated with a poor prognosis, and is also linked to an earlier age of onset (Eaton, 1991). Therefore it is of great significance to study pre-morbid adjustment in schizophrenics, with a view to increased accuracy of identification of those at risk.

The age of onset of schizophrenia is usually for convenience taken to be the time of first hospitalisation, which usually correlates with the first onset of positive symptoms (deLisi, 1992). However there is normally an insidious onset of negative symptoms before this, and subtle signs of abnormality are usually present since childhood. Two large prospective studies of British birth cohorts have attempted to correlate particular childhood behaviours with a greater risk for developing schizophrenia. Jones *et al* (1994) found that pre-schizophrenics reached milestones of motor development (especially walking) later, and had more speech problems. They had low educational test scores at ages 8, 11 and 15 and exhibited a solitary play preference at ages 4 and 6. In early adolescence, pre-schizophrenics rated themselves as less socially confident, and teachers rated them as more anxious in social situations. There was also a preponderance of twitches, grimaces and severe nail biting in pre-schizophrenics.

Similar results were found in a study by Done *et al* (1994). At age seven, they found a significantly greater degree of social maladjustment in pre-

schizophrenics, which was particularly significant in boys. These boys were rated as 'over-reactive', that is more anxious for acceptance and more deviant than their classmates, and displaying hostility and inconsequential behaviour. Social maladjustment in pre-schizophrenic girls however was not generally not noticeable until age 11. Signs of 'under-reactive' behaviour were then noted, such as timidity, withdrawal, lethargy, apathy and depression, suspiciousness and unwillingness to work.

A study of neonates born to schizophrenic mothers (therefore at high risk for the later development of schizophrenia) revealed an increase in abnormal neurological signs compared with controls (Blennow and McNeil, 1991). These neonates exhibited unusual wakefulness and displayed deviations in sensitivity to stimulation and tests of reflexes, complex responses and excitability/depression.

By assessment on the Waldrop Physical Anomalies scale, schizophrenics have been found to have an increased frequency of minor physical abnormalities (MPAs) (Lohr and Flynn, 1993). Such abnormalities include low set ears, a high steepled palate, a curved fifth finger and partial syndactyly between the two middle toes. Presence of MPAs in schizophrenics has been correlated with poor premorbid adjustment and early age of onset (Lohr and Flynn, 1993). The presence of such minor physical anomalies is normally indicative of developmental disturbance, and is therefore consistent with the neurodevelopmental hypothesis of schizophrenia (see section 1.7.2).

The occurrence of hand morphology discrepancies is of particular relevance to schizophrenia, since the ectodermal cells which will form the hand skin migrate in the fourth or fifth month of gestation, the same time as the bulk of neuronal migration to the cortex (Bracha, 1991). While under obvious genetic programming, this implies that hand morphology abnormalities can exist as

'fossilised' evidence for a prenatal insult such as ischaemia or toxicity. In a study of 24 pairs of monozygotic twins, it was consistently found that the schizophrenic co-twin exhibited more hand morphology abnormalities than the unaffected twin (Bracha, 1991). There was also a trend towards more dysmorphology of the right hand, which suggests a developmental brain lesion in the left hemisphere. This is consistent with some other observations of neuropathological changes in schizophrenics (see section 1.4.11).

1.4.10 Markers of predisposition

Since there is considerable evidence that schizophrenia occurs as a result of a neurodevelopmental lesion, it is theoretically possible to detect those at risk for the disorder well before the onset of the diagnostic symptoms. Since the precise neurodevelopmental anomaly is not yet known, work has focused on identifying genetically transmitted markers which are correlated with and specific to schizophrenia. Six criteria have been proposed for a good candidate marker (Holzman, 1992). It would have a low base rate in the population, would be stable over time and relatively specific to the diagnostic entity. It should also aggregate in the relatives of affected but not unaffected individuals, be genetically transmitted and identify those at risk for developing schizophrenia. Several traits have been identified which are good candidates for biological markers of schizophrenia. These are smooth pursuit eye tracking, event related potentials and continuous performance test ability.

Smooth pursuit eye tracking: When visually tracking a sinusoidally moving object, the eye will normally move smoothly, with occasional compensatory saccades when the eye moves too fast or too slowly. It has been found that an abnormality of eye tracking is associated with schizophrenia, such that 'catch-up' saccades are required more frequently and the smooth pursuit is interrupted by more random saccadic intrusions (Levy *et al*, 1994). Eye tracking dysfunction is found in 5-9% of normal individuals, 21% of non-

psychotic psychiatric patients and 53-86% of schizophrenics. A great number of studies have replicated these findings (reviewed in Levy *et al*, 1994).

Eye tracking dysfunction seems to be relatively specific to schizophrenia, and is not an effect caused by drug treatment. Importantly, the same deficit can often be seen aggregating in the non-psychotic biological relatives of schizophrenics, which suggests that it is genetically transmitted. Holzman (1974) found eye tracking dysfunction in 45% of clinically well first degree relatives of schizophrenics, as compared with 10.5% of relatives of non-schizophrenics and 8.3% of normal individuals. In addition, the concordance for eye tracking dysfunction between twins has been found to be twice as high in monozygotic pairs than dizygotic pairs, as would be expected for a genetic trait. The above findings fulfil all the criteria for a good marker except the last, which states that the marker should be able to identify those at risk. Eye tracking dysfunction is incapable of this because not all schizophrenics display the marker (14-50% of schizophrenics exhibit normal eye tracking), and not all the individuals with the deficit develop schizophrenia.

These observations led Matthysse (1986) to propose a 'latent trait' model for schizophrenia. They assumed that schizophrenia and eye tracking dysfunction are independent phenotypic expressions of a latent trait, transmitted by a single gene. This does not rule out the possibility that other genes affect the expressivity of the latent trait, and as such is not incompatible with the observed non-Mendelian segregation of schizophrenia in families. Experiments testing mathematical models of the structure and parameters of the latent trait hypothesis have been tested, have generally yielded results which closely fit those predicted (Levy *et al*, 1994).

Event related potentials (ERPs): ERPs are a measure of brain electrical activity related to sensory, motor and cognitive responses to specific stimuli. The P300, which reflects the cognitive work done by the subject after a

stimulus, has been found to be of a decreased amplitude in some schizophrenics (Holzman, 1992). There is probably also a small but significant delay in P300 latency, though this has not been proven by all studies undertaken. Well siblings of schizophrenics have also been found to show changes in the P300 component of the ERP.

Another component of the ERP is the P50 wave, which occurs as a response to repeated auditory stimuli. In most individuals, the magnitude of the P50 response diminishes considerably with repetition of the stimulus, due to habituation (Holzman, 1992). In schizophrenics however, the diminution of the P50 is much smaller. Adler *et al* (1982) found that in normals, the magnitude of the P50 after a second stimulus was only 13.9% of that after the first. In unmedicated, acutely psychotic schizophrenics however, the ratio was 90%. This effect has also been shown to occur in unaffected family members of schizophrenics and is therefore genetically transmitted.

Continuous performance test: There are many tests of continuous attention; in general they involve a series of stimuli presented to the subject briefly, and in quick succession. The subject must respond each time a target stimulus appears. In such tests, it has been found that 40% of schizophrenics showed a deficit in performance, with an increased frequency of errors of omission (not responding to the target stimulus) and commission (responding to a non-target stimulus) (Holzman, 1992). The effect was present both during acutely psychotic episodes and in remission, and was also found in unmedicated schizophrenics. It has also been detected in unaffected siblings and offspring of schizophrenics, and those diagnosed with schizophrenia spectrum disorders (Holzman , 1992).

1.4.11 Neuropathology

Structural Changes

The techniques of computed tomography (CT) and magnetic resonance imaging (MRI) can be used for brain imaging studies on living subjects. CT uses x-rays to differentiate between bone, brain tissue and cerebrospinal fluid (CSF). MRI has now largely superseded CT due to its greater resolution and no requirement for exposure to large doses of ionising radiation (Potts *et al*, 1993).

The first report of structural changes in the brains of schizophrenics was a CT study (Johnstone *et al*, 1976) showing that the ventricular spaces of schizophrenics were larger than those of controls. Since then there have been a great number of CT and MRI studies, from which the consistent findings are ventricular enlargement and cortical atrophy (Bachneff, 1991). There is also a localised volume reduction, with the temporal lobe, especially the hippocampus and the parahippocampal gyrus, being primarily affected (Roberts and Bruton, 1990). The pathology is usually of equal severity in both male and female patients and also between the left and right hemispheres although several studies have reported a greater degree of abnormality in the left hemisphere (Waddington, 1993). In addition, disturbed cytoarchitecture of the pyramidal cells of the hippocampus has been reported, and post mortem studies of schizophrenics have revealed a decrease in brain weight (average 6%) and brain length (average 4%) relative to controls (Roberts and Bruton,1993).

The nature of the structural pathology seen in the brains of schizophrenics is unfortunately not completely specific to the disorder. Structural brain changes are not seen in all schizophrenics. In one study of ventricular size 50% of the patient sample overlapped with the normal range (Cannon and Marco, 1994). However, specificity to pathology can be seen in twin studies, where the schizophrenic co-twin is always found to have larger ventricles

than the well co-twin (Reveley *et al*, 1992). These well co-twins do often have enlarged ventricles when compared with unrelated controls however. Increased ventricular size is also seen in other pathologies, such as Alzheimer's disease and Parkinson's disease. Hence it cannot be used alone as a diagnostic marker for schizophrenia.

There is also little evidence for subgroupings of schizophrenia being related to particular structural pathologies. An association between ventricular enlargement and negative symptoms has been proposed, but the evidence for this is inconclusive (seven positive out of twenty studies reviewed) (Cannon, 1991). It was also suggested that only sporadic cases of schizophrenia would show brain pathology, but evidence contrary to this argument has been supplied by family studies.

Functional changes

Functional studies of the brain have mainly relied on the technique of positron emission tomography (PET), which uses radioactive ligands to measure metabolic or neurochemical activity. The resolution of PET is sufficient to be able to study small structures, but use of the technique is severely limited by exposure of the subject to non-trivial amounts of ionising radiation. Non-radioactive techniques such as magnetoencephalography (MEG) and magnetic resonance spectroscopy (MRS) are safer but are not so widely used.

With the exception of the many PET studies on neurotransmitter levels and receptor densities (see section 1.7.1), the major findings of functional studies in schizophrenics are of hypofrontality (Cleghorn *et al*, 1991). Hypofrontality is a reduction in the hyperfrontal pattern which is seen in well individuals, as measured by the increase in blood flow and glucose utilisation when subjects are asked to perform tasks requiring use of the frontal cortex (e.g. the Wisconsin card sorting task). Hyperfrontality is known to be reduced only

slightly by the use of neuroleptic drugs.

1.4.12 Neurochemistry

The clinical manifestation of schizophrenia gives reason to suppose that there may be abnormalities in neurotransmission, either as a causative factor or at least as a reliable marker of the disorder. Consequently there is a plethora of experimental data both supporting or refuting the idea that neurotransmitters are involved in the disease. Unfortunately many of the earlier studies were plagued by methodological problems and errors and as such a clear picture is yet to emerge. Standardised experimental procedures in combination with the improved structural and functional brain imaging techniques now available should clarify the situation. The involvement of neurotransmitters in schizophrenia has been reviewed in section 1.7.1.

1.5 Treatment of Schizophrenia

Treatment for schizophrenia in the early part of this century and earlier was for the most part ineffective and often highly dangerous. Injection of adrenaline, fever induction and vasectomy proved of little therapeutic use (Hegarty 1994), whereas alternatives such as psychosurgery and sleep therapy carried tremendous risks (Frankenburg, 1994). In the 1920's Henry Cotton 'removed organs in a frantic desire to improve his cure rate, and probably only alleviated psychosis by terminating his patients' lives!' (Frankenburg, 1994).

The first effective treatments were not seen until the introduction of insulin coma, chemoconvulsive therapy and electroconvulsive therapy (ECT) in the 1930's. The first antipsychotic drug, reserpine, was derived from the roots of the plant *Rauwolfia serpentina*, known in central Asia for centuries as 'insanity herb' (Frankenburg, 1994). At the same time in the 1950's, chlorpromazine, developed as a surgical anaesthetic, was found to have

antipsychotic properties (Frankenburg, 1994). Reserpine and chlorpromazine became the prototype antipsychotic drugs, upon which many others have been modelled. These drugs are known as 'typical' neuroleptics.

These drugs are however far from perfect. They show an incomplete clinical efficiency, especially for the treatment of the negative symptoms, and can produce a disturbing array of extrapyramidal side effects (Livingston, 1994). The side effects of chlorpromazine for instance are undue sedation, autonomic symptoms, weight gain and movement disorders including drug-induced Parkinsonism and even tardive dyskinesia in 25% of chronically treated patients (Livingston, 1994). Indeed, the term 'neuroleptic', by which these drugs are now known, implies induction of extrapyramidal side effects in addition to antipsychotic properties.

The last decade has seen the development of 'atypical' neuroleptics - so called because they have antipsychotic properties but do not induce extrapyramidal side effects. The prototype atypical neuroleptic clozapine was withdrawn in the 1970's after it caused fatal cases of agranulocytosis, but was reintroduced in the 1980's and is now one of the most effective antipsychotics. It exhibits an enhanced efficiency in the alleviation of positive but especially negative symptoms compared to typical neuroleptics and has the lowest risk for inducing tardive dyskinesia (Pickar, 1995). Clozapine is also effective in up to 70% of previously treatment resistant cases.

1.5.1 Mechanism of action

Typical neuroleptics cause a blockade of neurotransmitter receptors, particularly the dopamine receptor D2. The specificity for D2 varies between different drugs, between the extremes of haloperidol (which blocks D2 and has little effect on any other receptors) and fluoxetine (which blocks serotonin reuptake only) (Gershon and Rieder, 1992). The atypical neuroleptics also block dopamine receptors, but do so specifically in the

mesolimbic dopamine system rather than the nigrostriatal system. This has led to the proposal that dopamine receptor blockade in the former is responsible for antipsychotic action, whereas blockade in the latter causes extrapyramidal side effects (Pickar, 1995). Clozapine is a weak D2 receptor blocker compared with conventional agents, and instead exhibits a higher efficiency for the D1 and especially the D4 receptors. It also has a high affinity for a broad range of serotonin receptors (especially 5HT6 and 5HT7), and has activity at adrenergic (α1 and α2), histaminergic (H1) and muscarinic receptors (Pickar, 1995).

The importance of considering the interactions between the different neurotransmitter systems, rather than focusing on dopamine, is becoming clear. The development of the latest antipsychotic drug risperidone was in response to the observation that addition of serotonin agonists such as ritanserin to the regimen of patients receiving haloperidol improved the treatment of negative symptoms, ameliorated depression and anxiety in schizophrenics and reduced movement disorder (Livingston, 1994). Risperidone is a highly potent antagonist of the serotonin receptor 5HT2, and is also active at D2 and α1 and α2 adrenergic receptors (Livingston, 1994).

1.5.2 Strategy for administration

The choice of drug used is determined by many factors such as the clinician's preference, side effects, the patient's medical history and the family history of response and tolerance, the availability of routes for drug administration and often financial considerations (Shen, 1994). Most patients are maintained on medication after discharge from hospital, which has been found to be 75% effective in relapse prevention compared with 20% relapse in individuals who received a placebo (Shen, 1994). The effectiveness of medication is obviously dependent on good compliance, which is often difficult to achieve due to lack of insight and side effects. The decreased side effects of atypical neuroleptics has improved compliance, although these

drugs must be administered orally as opposed to intramuscular injection (Hale, 1993).

1.6 Aetiology of schizophrenia

1.6.1 Evidence for a genetic component

Family studies

There have been many studies investigating the presence of schizophrenia and other psychiatric disorders in the relatives of schizophrenic probands. All except two (Pope *et al*, 1982 and Abrams and Taylor, 1983, which can be severely criticised on methodological grounds) agree that rates of schizophrenia are greatly increased in the relatives of schizophrenic probands. Consistent with a genetic effect, the risk increases the closer the relative, and hence the higher proportion of genes shared. Aggregated rates of schizophrenia in relatives gathered from forty European studies conducted between 1920 and 1987 suggest that siblings of schizophrenics have an age corrected risk of 9% of developing the disease, parents have a risk of 6%, offspring 13% and secondary relatives 2% (Prescott and Gottesman, 1993). Spouses have a risk of 2%, a figure which is greater than the general population risk of 1%, perhaps due to assortative mating. There is also elevated risk for schizophrenia spectrum disorders in families, and when these are taken into account a psychiatric morbidity may be found in approximately 35% of the relatives of schizophrenics compared with only 10% of controls.

These figures provide evidence that schizophrenia is not caused by a single gene with complete penetrance, since for such monogenic disorders the familial risk is directly proportional to the genetic similarity (50% in first degree relatives, 25% in second degree relatives etcetera).

Twin Studies

Studies of concordance for psychiatric disorder within pairs of monozygotic versus dizygotic twins are a useful way to show genetic influence, yet due to the relatively rare incidence of both twinning and schizophrenia, it is not easy to generate large numbers for credible studies. Most twin studies have therefore been conducted in Scandinavia where complete Psychiatric and Twin registers facilitate the systematic ascertainment of probands for a region or entire country (Fischer *et al*, 1969).

Widely varying concordance rates for schizophrenia between twins have been reported (Fischer *et al*, 1969), yet the salient point is that the concordance between monozygotic (MZ) twins is always at least twice as high as concordance between dizygotic (DZ) twins. This is consistent with a genetic model for schizophrenia, since MZ twins are genetically identical, yet DZ twins share only half their genes. In a review of eleven twin studies, Gottesman and Shields (1982) found an average concordance rate of 57.7% for monozygotic twins compared with only 12.8% for dizygotic twins.

Of concern to advocates of a 'purely genetic' aetiology is that concordance between monozygotic twins does not reach 100%. However risk of schizophrenia in the offspring of both twins from a discordant MZ pair is found to be the same, approximately 17% (Fischer, 1971). For discordant DZ twins the risk to the offspring of the schizophrenic twin is also 17%, yet the risk to the offspring of the normal twin is only 2%, not significantly different from the population as a whole (Fischer, 1971). This suggests that the normal twin of a discordant MZ pair does possess all the factors necessary for the development of schizophrenia, since they can be passed on to the next generation, but for some reason the phenotype is not expressed. This has been attributed to either incomplete penetrance of the genes involved, the necessity of a precipitating environmental factor, or both.

A more stringent test of a genetic aetiology of schizophrenia comes from studies of MZ twins reared apart, where one is known to have schizophrenia. This has an advantage over conventional twin studies since the effect of shared rearing environment (with the exception of the prenatal environment) is removed. Acquiring these subjects is obviously extremely difficult and to date there have been only sixteen pairs studied (Shields and Slater, 1967). Of these ten were concordant and six discordant for schizophrenia. This results in a concordance rate of 62.5%, which is similar to rates for twins reared together and suggests little or no effect of shared environment.

Adoption Studies

Adoption studies have an advantage over family and twin studies because it is possible to control for the effects of environment. There are two fundamental ways to approach them - either the 'adoptees study method' in which rates of the disorder are assessed in adoptees born to schizophrenic mothers, or the 'adoptees' relatives study method' whereby diagnosis of schizophrenia in the biological parents of schizophrenic adoptees is assessed (Tienari, 1994).

Whichever method is used, the results always indicate a genetic aetiology for schizophrenia. By the adoptees study method, Heston (1966) found that five out of forty-seven (16.6%) adoptees born to schizophrenic mothers developed the disorder themselves, compared with none of the fifty controls. However half of the group of index adoptees exhibited major psychosocial disorder. Similarly Rosenthal *et al* (1971) found a rate of 31.6% of schizophrenia spectrum disorders in the index cases compared with 17.8% in controls. In this study the mothers had not developed schizophrenia before the adoption took place, to ensure that the foster parents had no idea that the adoptee had an unusually high risk of becoming schizophrenic. Using the adoptees' relatives method, Kety *et al* (1975) produced similar results - 8.7%

of the biological parents of schizophrenic adoptees had psychiatric morbidity, compared with only 1.9% of the controls.

The latter study included an important subsample: the paternal half siblings of the adopted individuals. Paternal half siblings are related through a common father, and therefore they do not share the same prenatal, perinatal and post natal environment. They do however share 25% of their genes. Kety found that the risk of schizophrenia was greater in the biological paternal half siblings of the schizophrenic adoptees than in the paternal half siblings of control adoptees. Kety *et al* (1978) believed that this data was "the most compelling evidence we have that genetic factors operate significantly in the transmission of schizophrenia."

1.6.2 Genetic Models

Despite convincing evidence that genetic factors are involved in the aetiology of schizophrenia, the genetic mechanism(s) is proving remarkably elusive. Part of the problem is that the mode of inheritance is unknown, since schizophrenia does not exhibit classical Mendelian transmission in families.

The simplest genetic model is the single major locus (SML) model (Prescott and Gottesman, 1993). The basic observation that large multiply affected families are very unusual seems to preclude the existence of a highly penetrant dominant allele, although modifications such as reduced penetrance, imprinting, anticipation, variable age of onset and phenocopies can be incorporated. A small number of isolated pockets have been reported in which schizophrenia seems to be transmitted by a single dominant allele (for instance Iceland and Northern Sweden), but in general single gene effects would seem to make only a modest contribution to schizophrenia risk.

An oligogenic or polygenic model seems more favourable in this instance. The multifactorial polygenic model (MFP) assumes a large number of genes

of small and equal effect (Prescott and Gottesman, 1993). The mixed model (Morton and McLean, 1974), which is currently favoured, incorporates a polygenic background with or without a major locus and with or without environmental effects.

1.6.3 Imprinting and Anticipation

There is currently much debate as to the possible rôle of genomic imprinting and anticipation in schizophrenia, as potential ways to reconcile the apparent complex mode of transmission with the existence of genes of major effect.

As yet, no evidence has been found to suggest that genomic imprinting may play a rôle in psychiatric illness (Asherson *et al*, 1994).

The term 'anticipation' was first used by Mott in 1910, meaning an increase in severity of illness across successive generations. Mott clearly recognised the existence of the phenomenon in psychiatric illness, when he referred to the 'the law of anticipation of the insane' (Asherson *et al*, 1994). There have been several modern studies also seeking to prove that schizophrenia becomes more severe in successive generations. Bassett and Honer (1994) studied three generations of siblines selected for autosomal dominant like inheritance. They demonstrated that patients in younger generations were more frequently hospitalised, had an earlier age of onset and increased severity of the disorder. It is possible that ascertainment biases alone are responsible for this finding. There may be preferential ascertainment of parents with late onset due to reduced fertility in those with early onset. There may be preferential ascertainment of younger generation patients with an early onset, since not all would have passed the risk period for developing the disorder. Certainly not all studies have demonstrated the presence of anticipation - Asherson *et al* (1994) concluded that anticipation was not occurring in their sample of multiplex families and O'Neil *et al* (1993) postulated that the effect was merely 'regression to the mean', meaning an

early onset in the offspring of parents with a late onset and vice versa.

A molecular mechanism of anticipation has recently been identified as an expansion of unstable trinucleotide repeat DNA sequences, in several neurodegenerative disorders including Huntington's disease, myotonic dystrophy, Fragile X syndrome, spino-cerebellar ataxia, and spinal and bulbar muscular atrophy (Petronis and Kennedy, 1995). Much interest has been generated in a search for unstable trinucleotide repeat sequences in schizophrenia. Recently, two studies have been published which demonstrate a positive association between triplet repeat expansions and schizophrenia. The first (O' Donovan *et al*, 1995) used the repeat expansion detection method (RED) to demonstrate a significantly increased repeat length in a sample of schizophrenics and bipolar patients relative to controls. The association was particularly significant for females. No correlation was found between repeat length and age of onset, severity or the presence of particular symptoms. The second report (Morris *et al*, 1995) also used RED to demonstrate longer repeat lengths in schizophrenics than normals, an association that was again highly significant for females. In this study however an association was seen between a large repeat expansion and an early age of onset.

An expansion of trinucleotide repeat sequences could form the basis for an explanation of discordance in monozygotic twins, if a particular threshold number of repeats is required for expression of the phenotype. Uneven expansion of the repeats between the co-twins in the somatic cells during development could result in only one twin of the two reaching the critical threshold value. The unaffected twin would have a number of repeats just below the threshold; further expansion in his or her offspring could explain the higher than expected prevalence in the offspring of unaffected twins. If the number of repeats correlates with severity of the disorder, then the disorder would tend to follow a milder course in the affected twin of a

discordant pair than in those of a concordant pair, since the discordant pair would, on average, have a fewer number of repeats than the concordant pair. Most data are consistent with this (Petronis and Kennedy, 1995).

1.6.4 Somatic Genetics

Genetics may play a role in the aetiology of schizophrenia without inherited factors. Schizophrenogenic mutations taking place in critical somatic cells during development may provide a genetic explanation for a disease which outwardly may seem of sporadic and therefore environmental origin. In addition to simple mutation, unstable DNA sequences and epigenetic factors such as inactivation may be involved.

1.6.5 Evidence for an environmental component

Although the evidence for a genetic component in schizophrenia is overwhelming, it is by no means suggestive that genetic influences are necessary *and* sufficient to cause schizophrenia in all cases. The discordance for schizophrenia seen in some pairs of monozygotic twins, the non-Mendelian transmission of genetic factors and the lack of success in identifying causative genetic loci all point towards there being an environmental component in the aetiology of schizophrenia. Indeed many cases of schizophrenia occur in the absence of a known family history of the disorder and as such appear caused primarily by environmental factors. Weiner (1985) has estimated that only 10% of schizophrenics have a family history of schizophrenia (although 25% have a family history of schizoid personality disorder). Few researchers would argue however, that schizophrenia is caused by the environment alone, instead preferring a multifactorial aetiology combining genetic predisposition and precipitating environmental events (the diathesis-stress model).

Winter Birth

There is a proven excess of schizophrenic births during the Winter months, which is not merely attributable to analytical methods (e.g. the age-incidence

effect). There are three alternative explanations for this effect (Pulver *et al*, 1992). 1) There is some seasonally varying factor which affects intra-uterine life and increases the risk for the development of schizophrenia in genetically predisposed individuals. This may be infectious disease, a low protein diet or pregnancy and birth complications. 2) Individuals genetically at risk for schizophrenia have some biological advantage which protects them from allergies or infections which are more prevalent during the Winter months and may otherwise have caused mortality. 3) Seasonal patterns of conception among the parents of schizophrenics. There is no evidence for the latter.

Studies of the season of birth of schizophrenics from the Southern hemisphere will prove useful. Since Winter occurs in the middle of the year in the Southern hemisphere, the results would not be confounded by an age-incidence effect. Preliminary results (Welham *et al*, 1995a and b) suggest that an excess of pre-schizophrenic births in the Winter and early Spring months may occur in the Southern hemisphere.

Social Class

There have been many studies investigating the social class of schizophrenics, from which the general conclusion is that schizophrenia occurs most frequently among the lowest social classes. This finding may be viewed in two different ways, according to either the 'social causation' (low social class causes schizophrenia) or the 'social selection' (presence of schizophrenia causes drift towards the lower classes) hypotheses (Kohn, 1976).

There are environmental factors sometimes associated with the lower classes which may causes schizophrenia, such as greater environmental and occupational hazards, increased prevalence of infectious agents, poor maternal and obstetric care and limited resources to deal with stress in addition to cognitive and personality factors. Perhaps the difficult conditions

of life and a self-perception of helplessness which can be imposed by low social class lead to a failure to cope with stress and precipitation of a schizophrenic episode in those genetically predisposed (Kohn, 1976). The socially debilitating nature of the disorder however does also implicate drift towards the lower classes. The familial nature of schizophrenia is such that drift of schizophrenic parents or grandparents towards the lower classes would also contribute to this effect.

City Residence

It is interesting to note that the frequently proposed emergence of schizophrenia as a disease entity was in the late 19th century, at a time when the first large scale urbanisation of the population was also taking place. Since then, many studies have found prevalence of schizophrenia to be correlated with urban residence. This may be interpreted as either cause or effect. Adverse environmental factors may aggregate in cities (for instance pollution, infectious agents) yet on the other hand schizophrenics may drift towards the inner city, where there is more likely to be cheap, single person accommodation, casual labour and a lack of a close-knit social structure from which a schizophrenic individual may feel excluded (Freeman, 1994). Another alternative is the 'social residue' hypothesis, which implies a selective loss of mentally healthy individuals from socially and environmentally undesirable areas such as the inner city, leaving an excess of the psychiatrically disturbed and socially incompetent (Freeman, 1994).

Stress

It is found that independent life events (that is, those not related to an individual's personality or behaviour) tend to cluster before the onset or a relapse of schizophrenia. However stress is usually interpreted to be only a determinant of the time of a schizophrenic episode and not causative in those not genetically predisposed to the disorder (Leff, 1992).

Diet

Although less thoroughly documented, some associations between schizophrenia and grain in the diet have been proposed, in the presence of genetic vulnerability. A positive correlation has been proposed between the incidence of schizophrenia and the consumption of grains (Hemmings, 1990). Case studies provide additional evidence, such as two patients whose psychoses lessened when they were put on a grain free diet, only to relapse again when gluten was added (Hemmings, 1990).

Trace elements may be involved in the aetiology of schizophrenia. There are clinical similarities between schizophrenia and 'manganic madness' caused by manganese poisoning (Donaldson, 1987); Dawson (1970) has proposed that the high lithium concentration in the drinking water of some parts of Texas may be responsible for the low levels of psychiatric admissions there; Foster (1988) generated a correlation coefficient of 0.58 between schizophrenia and areas of the United States in which there was a low selenium level in the diet.

An increase in pre-schizophrenic births following pre-natal exposure to the Dutch hunger Winter of 1944-45 has been reported. The nutritional deficiency imposed during this famine is probably the cause, although other factors such as toxicity from consumption of unusual foodstuffs (such as tulip bulbs!) may be taken into account (Susser *et al*, 1995). Prenatal nutritional deficiency can alter the functioning of neurotransmitter systems and can affect the morphology and electrophysiological properties of the hippocampus (Butler *et al*, 1994).

Prenatal exposure to alcohol may increase susceptibility to the development of schizophrenia (Lohr and Bracha, 1989). A peak in beer sales in the late Summer has been linked with the excess of pre-schizophrenic Winter births, and the relatively high rate of alcoholism in Ireland may be responsible for

the high incidence of schizophrenia there. Foetal alcohol syndrome (FAS) has some clinical similarities to schizophrenia, and Alcohol Hallucinosis, which develops 15-20 years after exposure to alcohol is virtually indistinguishable from schizophrenia (Lohr and Bracha, 1989). Prenatal exposure to alcohol in rats has been shown to cause alterations in the organisation and volume of the hippocampus and also more general neuronal loss.

Viral Infection

It has been proposed that schizophrenia is a genetic 'morphism', with both favourable and unfavourable effects. It is the balance between the disadvantageous schizophrenic symptoms (and decreased reproductive fitness) and other, advantageous, properties conferred by the morphic gene which is responsible for the gene's maintenance at a relatively high population frequency (Huxley *et al*, 1964). The advantages conferred by this morphic gene may include an increased resistance to surgical shock, burns, arthritis and substances such as histamine and insulin. Thirty years ago, Eliot Slater put forward the proposal that the reason schizophrenia was not progressively eliminated from the population was due to an increased resistance to infection (Wright *et al*, 1993). Carter and Watts (1971) examined the frequency of commonplace infections in the first degree relatives of schizophrenics. Their results showed that the relatives of schizophrenics were just as likely as the controls to have suffered a bacterial infection, but they were significantly less likely to have been recorded as having a viral infection.

Since then there have been many studies investigating the proposed association between prenatal exposure to influenza virus and subsequent increased risk of developing schizophrenia. Many have focused on assessing the rate of development of schizophrenia in individuals who were *in utero* during the 'flu pandemic of 1957. Studies from Finland (Mednick *et*

al, 1988), England and Wales (O'Callaghan *et al*, 1991), Japan (Kunugi *et al*, 1992), Ireland (Waddington, 1992) and Australia (McGrath, 1994) have all reported a significant increase in the rate of schizophrenia. A study from Scotland (Kendell and Kemp, 1989) gave equivocal results, although upon reanalysis of the data it did appear to demonstrate a positive association. Moreover the effect has been shown to be specific to those exposed during the second trimester of gestation, between the 5th and 7th months, and in some cases was only statistically significant for female schizophrenics.

Not all studies however have confirmed the association. Studies from the USA (Torrey *et al*, 1991), the Netherlands (Selten and Slaets, 1994) and England (Crow and Done, 1992) have reported no increase in schizophrenic births following exposure to influenza virus *in utero*. In the pandemic studies there is no proof that the mothers of later schizophrenics were actually infected. Crow and Done (1992) found no association between 'flu and schizophrenia when mothers' reports of 'flu symptoms were taken into account. The method of data collection in this study is however not wholly reliable (midwife interview of mother and study of case notes) and it must also be borne in mind that up to 30% of people could have been infected subclinically.

1.6.6 Environmental and/or genetic factors

Obstetric Complications

An excess of pregnancy, obstetric and peri-natal complications has been found among schizophrenics, particularly males (Cantor-Graae *et al*, 1994). Obstetric complications have also been equivocally linked with Winter birth, and a negative family history of schizophrenia (Cantor-Graae *et al*, 1994). Consistent with the neurodevelopmental hypothesis of schizophrenia it is presumed that these difficulties cause damage to the foetal brain, for instance by anoxia or haemorrhage. It is known that birth complications can result in enlarged cerebral ventricles.

Birth weight

Predisposition to later schizophrenia has been found to be associated with low birth weight. In monozygotic twins discordant for schizophrenia it is usually the affected co-twin who was lighter at birth (Rifkin *et al*, 1994). In the general population, low birth weight is associated with reading difficulties, language disorders and lower IQ, it is therefore possible that the low birth weight occurs secondary to some form of neurodevelopmental pathology (Rifkin *et al*, 1994). In the presence of other genetic and perhaps environmental factors, this may be manifested as schizophrenia.

Head circumference

Some studies have found an association of schizophrenia with a small head circumference at birth (Kunugi *et al*, 1995 and McNeil *et al*, 1993). Others however have found no association (Kunugi *et al*, 1994).

Ethnicity

A British study of schizophrenics' racial background found that Afro-Caribbeans had higher rates of admission to psychiatric hospitals than did Caucasians, and more were diagnosed with schizophrenia (King *et al*, 1994). In general, Afro-Caribbeans were found to have a twelve fold increase in risk for developing schizophrenia relative to Caucasians and generally earlier ages of onset (King *et al*, 1994). This may be due to biological factors, such as differences in HLA antigens, but may also have a social component such as the effect of emigration and experience of racism.

1.7 Hypotheses on the aetiology of schizophrenia

1.7.1 An abnormality in neurotransmission

Dopamine

The origins of the dopamine hypothesis of schizophrenia were based on two primary observations. The first is that dopamine agonists such as amphetamines can produce a psychosis similar to schizophrenia in normal

individuals, and conversely drugs that potentiate dopamine activity generally worsen the symptoms of schizophrenia. Secondly, the primary action of neuroleptic drugs was found to be a blockade of dopamine receptors (Davis *et al*, 1991). Taken together, a hyperdopaminergic hypothesis for schizophrenia was proposed.

The evidence used to formulate the hyperdopaminergic theory came from studies investigating the concentration of dopamine and its primary metabolite homovanillic acid (HVA) in plasma and cerebrospinal fluid (CSF), and the density of dopamine receptors in the brain. Studies investigating the concentration of HVA in the CSF have yielded conflicting results, which can probably be attributed to methodological problems and intra-individual variation (Davis *et al*, 1991). Post mortem studies of dopamine and HVA concentration in schizophrenic brains consistently show differences from controls (usually an increase) (Davis *et al*, 1991), yet variability is seen in the anatomical site exhibiting the difference. These specificities may be a genuine finding, yet they may also be merely a product of varying analytical and statistical methods, a medication effect or a consequence of schizophrenia.

There are five types of postsynaptic dopamine receptor, termed D1-D5. The D1 receptor (Sunahara *et al*, 1990), is found primarily in the cortex, and is known to be positively coupled with adenylate cyclase, leading to an increase in cAMP. The D2 receptor on the other hand (Grandy *et al*, 1989) is found in striatal and limbic areas and leads to a decrease in cAMP. D2 has two subtypes, D2a and D2b, which cannot as yet be differentiated pharmacologically or physiologically. The D3 receptor (Sokoloff *et al*, 1990) is similar to the D2 but its expression is restricted to the limbic areas. The D4 receptor (Van Tol *et al*, 1991) has a particularly high affinity for clozapine. D5 (Sunahara *et al*, 1991) is similar to D1 but with a higher affinity for dopamine.

Post mortem studies of brain D2 receptor concentration have fairly consistently shown an increase in D2 receptor density in schizophrenics (Owen and Simpson, 1994). This finding has been seen in drug-treated patients but also in drug-naïve patients, which suggests that it is not merely an effect of neuroleptic medication. Recently Seeman *et al* (1994a) have found a 6-fold increase in levels of D4 in schizophrenic brains relative to controls.

There are a number of problems with the simple hyperdopaminergic theory, particularly the conflicting results with respect to plasma and CSF HVA concentration and the incomplete action of neuroleptics. An attractive extension to the dopamine hypothesis has been proposed (Davis *et al*, 1991), which takes into account observations essentially ignored by the original theory. These are that damage to the frontal lobe in animals leads to a deficit state similar to the negative symptoms seen in schizophrenia, and also that hypofrontality has been clearly demonstrated in schizophrenics. This hypofrontality appears to be associated with the dopamine system, since administration of amphetamines to schizophrenics causes an increase in prefrontal blood flow and a concomitant improvement in performance on the Wisconsin card sorting task. In addition, schizophrenia has clinical similarities with Parkinson's disease, which is known to involve a loss of dopaminergic neurons.

The model (Davis *et al*, 1991) proposes that schizophrenia is associated with *hypo*dopaminergic function in the frontal lobes, which can lead to *hyper*dopaminergia in the mesolimbic system. The former would explain the presence of negative symptoms, (untreatable by neuroleptics which decrease dopamine function) and could also explain the lower plasma and CSF HVA levels sometimes seen in schizophrenics. The latter would be characterised by the positive symptoms. If correct, this theory has implications for treatment strategies, since it is possible that premorbid amelioration of the cortical

hypodopaminergia with D1 or D5 agonists (since these receptor subtypes are most prevalent in the cortex) may prevent the later onset of psychosis.

Serotonin

Evidence for involvement of serotonin (5HT) is derived through similar channels as that for dopamine, namely that serotonin agonists (such as the hallucinogenic drug lysergic acid diethylamide, LSD) can precipitate a psychotic episode in a normal individual, and that serotonin antagonists are effective in the treatment of schizophrenia (Owen and Simpson, 1994). Many neuroleptic drugs have a high affinity for serotonin receptors in addition to dopamine receptors, and those that do are often particularly effective in treating the negative symptoms (for instance setoperone and ritanserin). It is possible that abnormalities in the serotonin system play a rôle in Type II schizophrenia, that characterised by negative symptoms (Bleich *et al*, 1988).

Studies on the serotonin system have mirrored those undertaken for dopamine. Plasma and CSF levels of the primary metabolite 5-hydroxyindoleacetic acid (5HIAA) have yielded inconsistent results, as have post mortem investigations of serotonin receptor density (Bleich *et al*, 1988). However, cortical atrophy and ventricular enlargement are irrefutably associated with a decrease in 5HIAA (Bleich *et al*, 1994), and some studies reveal an increase in 5HT in platelets from schizophrenics. This may be a consequence of a decrease in levels of monoamine oxidase B (MAO-B) which is the principal degradative enzyme for catecholamines (Mallet *et al*, 1994).

Studies on the concentrations of serotonin receptors (of which fifteen have been cloned) are in their infancy, but there is a suggestion of deceased levels of 5HT2A and increased levels of 5HT1a in schizophrenic brains (Roth, 1994).

Glutamate

Glutamate is the primary excitatory neurotransmitter in the brain, with a ubiquitous distribution (Owen and Simpson, 1994). Evidence for its involvement

in schizophrenia comes from the observation that phencyclidine (PCP), a potent glutamate antagonist, can cause a schizophrenia-like psychosis. In many ways this psychosis is a better model for schizophrenia than are those induced by dopamine or serotonin agonists, since this is characterised by both positive and negative symptoms (Owen and Simpson, 1994).

Glutamate receptors are divided into two types, NMDA (N-methyl D-aspartate) and non-NMDA. The former has six subtypes and the latter may be split into (R,S)-α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA)-preferring or kainate preferring (Szpirer *et al*, 1994). Experiments have revealed a decrease in non-NMDA receptors (Kerwin and Harrison, 1995) and an increase in NMDA receptors (Beckwith *et al*, 1995) in schizophrenics, also an upregulation of NMDA receptor mRNA in the hippocampus and a decrease in CSF glutamate concentration. This indicates a general hypofunction in the glutamate system in schizophrenia, which is consistent with the neurodevelopmental hypothesis (section 1.7.2), since excitatory amino acids are important trophic factors in the development of the hippocampus.

There are two alternative explanations for the mechanism by which glutamate may cause disrupted development. Firstly, glutamate is involved in mediating excitotoxity. Administration of glutamate causes opening of sodium channels and subsequent excitation, but also leads to an influx of water which causes toxicity and necrosis (Brown, 1992). Glutamate can also lead to apoptosis by causing an influx of calcium and subsequent increase in superoxide. The second theory relies on the fact that glutamate influences the organisation and stability of microtubules, via phosphorylative modification of microtubule associated proteins (MAPs). It has been hypothesised (Kerwin, 1993) that excessive release of glutamate during development may lead to disruption of microtubule assembly and a subsequent developmental anomaly in the temporal lobe, causing schizophrenia.

γ-amino butyric acid (GABA)

GABA is a major inhibitory neurotransmitter. For decades, convulsive therapy with metrazol (a GABA antagonist) has been used to treat schizophrenia, and many antipsychotic drugs act to reverse GABA's inhibitory effect (Squires and Saederup, 1991). This suggests a rôle for an excess of GABA in the aetiology of the disorder. Some studies however have reported a decrease in glutamic acid decarboxylase (the enzyme which converts glutamate into GABA) in schizophrenics (Owen and Simpson, 1994), which suggests the opposite.

Acetylcholine (ACh)

Nicotinic acetylcholine receptors are implicated in the aetiology of schizophrenia due to the association of smoking behaviour with schizophrenics (deLeon *et al*, 1995) and also because administration of nicotine to schizophrenics has been found to transiently normalise the deficit in auditory sensory gating (Leonard *et al*, 1995). A very high percentage of schizophrenic patients smoke (Lohr and Flynn, 1992); the nicotine may reduce negative symptoms, reduce extrapyramidal side effects and reduce the level of neuroleptic drugs in the blood, hence smokers require higher doses. The α7-nicotinic acid receptor has been cloned and found to be decreased in the hippocampus of schizophrenics (Leonard *et al*, 1995).

Noradrenaline

It has been proposed that the lack of goal-directed behaviour seen in schizophrenia could be the result of a deficit in noradrenergic function, and the authors subsequently reported a drop in dopamine β-hydroxylase (the enzyme catalysing the last step in the synthesis of noradrenaline) activity in schizophrenics (Owen and Simpson, 1994). However this has not been replicated and it is possible that the finding was a result of drug treatment.

Cholecystokinin (CCK)

There has been found to be a decrease in CCK mRNA in the temporal cortex and frontal lobes of schizophrenics relative to controls (Humphries *et al*, 1995).

Some post mortem studies have found a decrease in CCK receptors, yet others have not (Owen and Simpson, 1994).

Opioids

The opioid β-endorphin has been reported to cause a muscular rigidity similar to that seen in the catatonia of schizophrenia, but studies of plasma and CSF levels give conflicting results. Injection of the opioid receptor antagonist naloxone is reported to cause a decrease in hallucinations (Owen and Simpson, 1994).

Prostaglandins

The prostaglandins, especially prostaglandin E, are known to modulate monoamine transmission. The activity of prostaglandin E may be decreased in schizophrenics, thus resulting in increased dopamine and noradrenaline transmission (Neylan and vanKammen, 1990). Abnormalities in the essential fatty acids (EFAs) which are precursors for prostaglandins, form the basis of the membrane hypothesis of schizophrenia (Horrobin and Huang, 1993).

Interactions between neurotransmitter systems

It is becoming clear that dysfunction in a single neurotransmitter system is inadequate to explain the aetiology of a disease as symptomatically complex as schizophrenia. It is also realised that the different neurotransmitters do not act alone, but rather the systems interact to give one integrated response. As such, the normal state is one of elasticity, where transient changes in one system can be constantly compensated for or amplified by another. It has been proposed that the pathology seen in psychiatric disorders is a result of diminished plasticity in the neurotransmitter systems.

1.7.2 A neurodevelopmental anomaly

There has been considerable debate on the origin of the neuropathology seen in schizophrenia. Kraepelin believed the disorder had the

characteristics of a neurodegenerative process and the results of early CT studies were used to substantiate this idea. However the absence of correlation between length of illness and degree of degeneration did not fit the hypothesis, neither did the visualisation of structural brain changes in newly diagnosed patients. This instead suggested that the pathology was of neurodevelopmental origin (Roberts and Bruton, 1990).

In favour of this is the noted absence of gliosis accompanying the neuropathology. Gliosis occurs as a result of neuronal damage and is characterised by a proliferation of glial cells with a concomitant rise in associated glial proteins such as glial fibrillary acidic protein (GFAP), and enzymes such as monoamine oxidase B (MAO-B) (Roberts and Bruton, 1990). A destructive lesion early in development (less than 6 months) would leave no trace of gliosis, yet the type of pathology seen is indicative of the last trimester of gestation and as such this is unlikely.

Other evidence for the neurodevelopmental hypothesis can be seen by the fact that pre-morbid changes are seen in preschizophrenics (section 1.4.9), in social behaviour and also on neurological and dermatoglyphic measures. The excess of obstetric and perinatal complications seen in schizophrenics may be as a result or be the causative factor of a neurodevelopmental anomaly.

Crow (1980b) put forward a hypothesis that schizophrenia occurs as a result of a mutation in a 'cerebral dominance' gene, or that which is responsible for determining the asymmetry in size of the temporal lobes in the brains of humans. Temporal lobe asymmetry is a relatively late evolutionary phenomenon, believed to be linked to the development of language skills in humans. This may be linked to handed-ness, which is believed to be determined by a single dominant gene, termed the 'cerebral dominance gene' or 'right shift factor'. Crow (1980b) proposed that the location of this

gene was within the pseudoautosomal region of the sex chromosomes (see Section 1.10.1). Evidence for this theory is that the neuropathology seen in schizophrenics is often more apparent on the left side of the brain. Also that when psychotic symptoms accompany temporal lobe epilepsy, they are schizophrenia-like if the focus is on the left hand side (Crow, 1980b). Finally in a post mortem study (Crow *et al*, 1989), neuropathology was seen on both sides of the brain in Alzheimer's patients, yet only on the left side in schizophrenic patients.

The maternal viral hypothesis

In agreement with the neurodevelopmental theory of schizophrenia, it is proposed that influenza may cause schizophrenia via an effect on the foetus. This effect may be mediated by an auto-immune mechanism. The complement of HLA alleles possessed by the mother determines her immune response to influenza infection, and in certain cases it is proposed that maternal antibodies produced against the virus will cross the immature foetal blood-brain barrier and cross react with brain proteins. They perhaps interfere with neuronal migration, which is known to occur during the second trimester.

Evidence for the hypothesis can be seen when one considers other genetic and environmental aspects known to be associated with schizophrenia. There have been many suggestions of an association between schizophrenia and the immune response. Schizophrenics are reported to have elevated levels of immunoglobulins and α-interferon, and alterations in cellular immunity (Wright *et al*, 1993). Linkage studies have proposed, among others, a positive association between paranoid schizophrenia and HLA-A9 (McGuffin and Stuart, 1986).

Some viruses, for instance influenza, Eppstein Barr virus and cytomegalovirus, have capsular neuraminidases which can alter the sialic

acid moieties on neural cell adhesion molecules (NCAMs) which regulate the migration of neurons during embryogenesis (Wright *et al*, 1993). It is known that in some cases antibodies directed against particular brain components can be induced by viral infection (by for instance Group A streptococci or Neisseria meningitides), and in the rabbit, anti-influenza antibodies are produced which can cross react with mammalian and human hippocampal, cortical and cerebellar tissue (Wright *et al*, 1993). Ironically, those mothers with the optimal immune response may be those most likely to produce auto-immunity in the foetus yet themselves would be least likely to display overt symptoms of influenza infection (Knight *et al*, 1987). Interestingly, autoimmune diseases are more frequently seen in females, so it is noteworthy that the influenza association with schizophrenia is possibly only valid for females. It appears that other viral infections apart from influenza are not involved in the pathogenesis of schizophrenia. O'Callaghan *et al* (1994) studied the effect of infection by 16 different viruses, and none was shown to be associated with an increase in incidence of schizophrenia.

A number of environmental observations on the disorder are also consistent with the maternal viral hypothesis. There is an excess of schizophrenic births during the winter and spring, coinciding with the time of most viral infections. In addition, schizophrenia is more prevalent in urban areas, where viral infection would spread most effectively. The association of obstetric complications with schizophrenia is interesting as this represents another time when the foetal blood brain barrier may be compromised (Wright *et al*, 1993). Males are known to be more prone to birth difficulties and perinatal complications, so it has been suggested that this aetiology explains the sporadic cases of schizophrenia in males, whereas virally-mediated autoimmunity is responsible in females.

Other prenatal disturbances such as the effect of certain dietary deficiencies or toxicity have been implicated in the aetiology of schizophrenia. Recently

an association between prenatal Vitamin A excess or deficiency and schizophrenia has been proposed (Goodman, 1995). This is due to the observations that 1) the genes involved in metabolism and regulation of Vitamin A and the retinoids lie in chromosomal contiguity with several schizophrenia candidate genes, 2) that the former genes are known or postulated to be involved in the regulation of the latter, and 3) that retinoid dysregulation can cause pathologies and cranio-facial dysmorphisms similar to those seen in schizophrenia.

1.7.3 The membrane hypothesis

Levels of certain essential fatty acids (EFAs) have been found to be reduced in the membranes of red blood cells of schizophrenics, which have a composition very similar to cells in the brain. The membrane hypothesis proposes the existence of a defective acyl transferase (which inserts a type of EFA into the phospholipid membrane) or an over active phospholipase A (which removes EFAs from the membrane) (Horrobin *et al*, 1994). Consistent with the latter are the facts that levels of phospholipase A have been found to be elevated in some schizophrenics (Gattaz, 1990) and neuroleptic drugs are known to inhibit phospholipase A. In addition, the activity of three erythrocyte membrane ATP-ases (Na-K, Ca and Ca-Mg) has been found to be reduced in schizophrenia and depression, which is linked to increased activity of phospholipases (Rybakowski and Lehmann, 1994). The membrane hypothesis can account for a number of epidemiological and pathological features of the disorder (Horrobin *et al*, 1994).

1.7.4 Schizophrenia as cerebral diabetes

The effects of smoking and exposure to heterocyclic amines affect components of the pathways of glucose metabolism, leading to a deficiency in uptake of glucose into the cells and a state akin to diabetes. Smoking behaviour is highly prevalent among schizophrenics; heterocyclic amines, which cause neurotoxic damage, are more commonly found in the inner city

where schizophrenia is also more commonly found. Nicotine and heterocyclic amines have opposing effects on dopamine transmission, which could perhaps resolve the discrepancy of findings suggesting either hyper- or hypo-dopaminergia in schizophrenia (Holden *et al*, 1994).

As regards a diabetic state, the incidence of diabetes is more than 50% in the first degree relatives of psychiatric patients which suggests some familial loading between the two disorders, also insulin coma therapy was one of the first effective treatments for schizophrenia. Thirdly, hypoglycaemia can cause an overactivation of NMDA receptors (Holden *et al*, 1994), although it is hypofunction of the glutamate system which is generally proposed.

1.8 Positional Cloning

Positional cloning, or 'reverse genetics' as it used to be known, is a method for locating genes by a sequential narrowing of the genomic region in which it may lie (Collins, 1992), made possible by the advent of recombinant DNA technology. In the last twenty years it has superseded the classical methods of 'functional cloning', in which the disease gene was approached from knowledge of the biochemical defect and protein involved.

1.8.1 Linkage Analysis

The first step in this process of positional cloning normally involves genetic linkage analysis. This is a general term encompassing methods used to assess the proximity of a particular marker to the disease locus by analysis of the frequency of recombination between the two. A marker very close to a disease gene locus will recombine with the disease locus very infrequently, whereas an unlinked marker will recombine with the disease gene locus at a rate indistinguishable from that of markers across the rest of the genome. The number of informative meioses limits the sensitivity of recombination analysis. One recombination between two markers in 100 meioses

corresponds to a genetic distance of 1cM between the markers, which in man translates to an average physical distance of approximately 1Mb, though this varies throughout the genome due to local recombination 'hot spots' and 'cold spots'.

1.8.2 Global mapping of the human genome

The isolation of markers is a critical step in positional cloning; genetic markers to enable more accurate linkage analysis and physical markers to facilitate finer scale mapping once the approximate location of the disease gene has been deduced.

In order to be useful in genetic linkage analysis, markers must contain several different forms or alleles, that is, they must be polymorphic. Around a disease gene locus, one allele from a tightly linked locus will tend to segregate with the disease phenotype in a population, such that the disease phenotype may be associated with a specific haplotype. This is termed linkage disequilibrium, and implies that the disease mutation has arisen on a specific chromosome and that there has been insufficient time since the appearance of the mutation for random reassortment of closely linked markers. The more polymorphic the marker, the more useful it is, since it will be informative in a higher proportion of cases. The value of a marker is thus estimated by its polymorphism information content (PIC) value.

The first polymorphisms to be useful in linkage studies were restriction fragment length polymorphisms (RFLPs). RFLPs occur where a polymorphic base pair results in the production or absence of a restriction enzyme site, which may be detected by Southern blotting. The first comprehensive RFLP maps of the entire human genome became available in 1987, with an average of 10-15cM between the markers but many unmapped gaps.

A problem with RFLPs is that most have only two alleles and as such the majority are uninformative in most families. Minisatellites (or Variable Number of Tandem Repeats, VNTRs) are more useful since they have more alleles. The differences between VNTR alleles lie in their length, which can be detected by digestion at a flanking restriction enzyme site and then Southern blotting. The initial reports were of the existence of approximately 100 VNTRs, and few more have been reported since then. Unfortunately VNTRs do not have an even distribution throughout the genome, but rather tend to be clustered towards the telomeres in human DNA (Hearne *et al*, 1992).

Since 1989, microsatellites have become the genetic markers of choice. Microsatellites, or simple sequence repeats (SSRs) consist of di-, tri-, tetra- and even penta-nucleotide repeat sequences of varying length, which are short enough to be detected by PCR using flanking sequence primers (Litt and Luty, 1989 and Weber *et al*, 1989). Microsatellites are randomly distributed in the genomes of all eukaryotic organisms examined, except yeast, at a frequency of approximately one every 30-40kb. This high frequency, and the fact that they can be detected and scored by a simple PCR reaction make them extremely useful for mapping purposes.

The implementation of PCR as a standard laboratory technique has revolutionised physical as well as genetic mapping of the human genome. PCR can be a quick and specific way to detect the presence of a marker sequence, and it is highly amenable to automation. Indeed most physical markers now available are short (200-500bp), specific, PCR-amplifiable sequences of DNA, termed 'Sequence Tagged Sites' (STSs). The concept of an STS map of the human genome was proposed six years ago (Olson *et al*, 1989) as a 'common language' for physical mapping. It was then envisaged that in 5 years, an 100kb-resolution STS map of the entire human genome could be in place. Although this was a little optimistic, the principle of STS

mapping was adopted worldwide. The first five year goal of the Human Genome Mapping Project (HGMP), as defined in 1990, included the production of high resolution genetic maps, with average spacing of 2cM, and physical maps consisting of yeast artificial chromosome (YAC) contigs with STS's spaced every 300kb (Jordan, 1992).

The pooling and standardisation of results and resources has greatly facilitated worldwide collaboration on the HGMP. Enormous numbers of markers are generated by large scale semi-automated projects, such as Généthon in France, and because they are named according to a standard protocol and information about them is publicly available through databases, the efforts of all mapping laboratories can be directed towards a unified goal. The success of standardisation has been reflected by the publication of linkage and physical maps of the entire human genome.

The first comprehensive linkage maps of the human genome were published in 1992 (NIH/CEPH collaborative mapping group, 1992). This map contained 1416 loci, consisting of RFLPs, protein polymorphisms, single base change polymorphisms and microsatellites. These markers had been typed onto a set of reference pedigrees made available by the Centre d'Étude du Polymorphisme Humain (CEPH), which maintains a regularly updated genotype database. An estimated 92% of the autosomal length of the human genome was spanned by the map.

Although these maps contributed considerably to the mapping of a number of simple monogenic genetic traits, the markers were not evenly distributed throughout the genome, and the density of markers was insufficient to be useful in the mapping of complex disorders or for refining linkage intervals to distances suitable for gene identification. Incorporation of more microsatellite markers was instrumental in the construction of more detailed subsequent maps (Weissenbach *et al*, 1992). The most recent linkage map of the human

genome (Gyapay *et al*, 1994) includes 2066 'CA' microsatellite markers, typed by linkage analysis on eight large CEPH families. The average distance between markers is now only 2.9cM, with 56% of the map covered by markers at 1cM resolution and only one gap of over 20cM remaining.

1.8.3 Fine mapping of regions of interest

Once the region of interest has been identified by marker analysis or a cytogenetic abnormality, it is often necessary to isolate more, densely spaced markers from that region, such that it may be thoroughly analysed.

To aid the positioning of new markers, a large scale restriction map of the region is useful. This can be achieved with pulse field gel electrophoresis (PFGE) (Schwartz and Cantor, 1984), which is capable of resolving fragments of more than a twelve megabases (Orbach *et al*, 1988), although it is disadvantaged by its time-consuming and technically difficult nature. Alternatively, fluorescent in situ hybridisation (FISH) could be used. Hybridisation of FISH probes onto metaphase chromosomes requires markers to be separated by more than ten megabases to achieve resolution, whereas hybridisation onto the less compact interphase chromatin can resolve markers separated by approximately 100kb. New techniques using stretched chromatin have increased resolving power still further, to almost 5kb.

Somatic cell hybrids can be a useful tool both for the generation of new markers and for the mapping of existing ones. They are made from a fusion of human and rodent cells, after which the rodent cells segregate the human chromosomes, resulting in a much reduced human content (Weiss and Green, 1967). The content can be further reduced by either exposing the cell to ionising radiation and then fusing it with a recipient cell line (radiation hybrids) (Goss and Harris, 1975), or chromosome mediated gene transfer (CMGT) (McBride and Ozer, 1973) in which a precipitate of donor fragments

and $CaPO_4$ is applied to recipient cells. The stable hybrids thus produced contain a small fraction of the human genome, which is useful for intermediate-scale mapping of markers. For instance a complete radiation hybrid map of 506 STS markers across chromosome 11 has been assembled (James *et al*, 1994).

A somatic cell hybrid which segregates a small specific part of the human genome can also be a useful tool for the isolation of markers from that region. Alu-PCR (Nelson *et al*, 1989) uses primers derived from the consensus of the Alu repeat to amplify inter-Alu fragments. Alu elements are specific to human DNA, thus Alu-PCR will generate markers from the human sequences in a somatic cell hybrid. Alu-PCR products can be used in bulk as a FISH probe, or individual products can be used as a probe or sequenced to create STS's.

The techniques of coincidence cloning (Devon *et al*, 1996) can give a higher degree of specificity in marker isolation. A hybrid which contains only the region of interest is rarely found; it is much simpler to define the region of interest as that portion of the human genome shared by two hybrids. One way to extract such sequences is to isolate the coincident sequences between a pool of Alu-PCR products from each hybrid by a selective cloning method (Aslanidis and deJong, 1991). The more stringent method of end ligation coincidence sequence cloning (EL-CSC) (see section 1.8.5 for a more detailed explanation) has been used to isolate single copy markers coincident between somatic cell hybrids and a pool of microdissection clones (Brookes *et al*, 1994).

On a larger scale, coincidence cloning using FISH technology can be used in the isolation of markers from a region shared between two chromosomes. The technique of 'coincidence painting' (Bailey *et al*, 1993) was designed for use between a derivative chromosome from a translocation cell line and flow

sorted normal chromosomes. The derivative chromosome is simplified, labelled and used as a FISH probe onto a metaphase spread of the normal chromosomes. After removing non-specifically bound probe by washing, the region of coincidence can be visualised, and the product eluted off and amplified. A similar method, in which the product is microdissected from the slide, is termed 'Prep-ISH' (Hozier *et al*, 1994) (see section 1.8.5).

Microdissection and microcloning of a region of interest can be used to generate markers. Microdissection of banded chromosomes (Senger, 1990) can yield abundant, short, region-specific probes. The microdissection technique used in this project was adapted from Weith (1994) which uses unstained chromosomes viewed under phase contrast microscopy. This yields clones of much greater size than the former technique. The validity of microdissection and microcloning to obtain markers can be seen in this project by the frequency of clones in close proximity to the translocation breakpoint.

Recent improvements in DNA sequencing technology have made the random sequencing of clones and subsequent generation of STSs a feasible way to obtain densely spaced markers. The concept of 'Genomic Sequence Sampling' (Smith *et al*, 1994) involves STS's across the whole genome, separated by a short PCR-able distance. The first step is the construction of highly redundant cosmid contigs across genomic regions, which would be subjected to fingerprint analysis to orient and order them, and then the ends of each cosmid would be sequenced to create a large set of ordered STS's. PCR between two adjacent STS's could give direct access to any part of the genome, and greatly facilitating fine mapping and sequencing to detect genes. The feasibility of genomic sequence sampling has been modelled in a simple eukaryotic organism (Smith *et al*, 1994).

Before the sequencing of the whole human genome begins in earnest, the priority of the HGMP is the establishment of sufficiently dense marker maps that any desired region of the genome can be accessed by PCR easily and quickly. An estimated 3% of the human genome only is believed to code for protein, and thus whole genome sequencing efforts were previously more usefully directed at simpler organisms with less redundancy in their genomes. In 1992 a project began to sequence the complete genome of the nematode *Caenorhabditis Elegans*, which is approximately 100Mb long (Sulston *et al*, 1992). Genes in *C.Elegans* are generally much smaller and have fewer and shorter introns than their mammalian counterparts, and the gene density is high. In addition, almost the entire genome was already available in ordered and cloned YAC and cosmid form. A follow up two years later reported the completion of sequencing for 2.2Mb of chromosome III (Wilson *et al*, 1994).

1.8.4 Contig Assembly

Once a set of closely spaced markers is available for a region, a contig of cloned recombinant DNA can be constructed. Many types of vector are now available, which vary in the amount of recombinant DNA which can be propagated in them and in their ease of use. The largest capacity cloning vectors commonly used are yeast artificial chromosomes (YACs) (Burke *et al*, 1987). More than one megabase of DNA can be cloned into YACs, making them the usual vector of choice for preliminary large scale cloning of a region. YACs are grown in the yeast *Saccharomyces cerevisiae* and so are easy to propagate, and their large size makes genetic manipulation straightforward. There are however several problems associated with YACs. They are difficult to isolate in a pure form due to contamination with yeast chromosomes, and have a low cloning efficiency. Deletions within the recombinant DNA are a fairly common occurrence, as is chimaerism resulting from co-cloning events or recombination between copies of the YAC within the yeast cell. For these reasons, greater reliance is now being

placed on PACs (P1 artificial chromosomes) (Ioannou *et al*, 1994) and BACs (Bacterial artificial chromosomes) (Shizuya *et al*, 1992). Approximately 100kb of human DNA can be cloned into PAC vectors, which also contain the ability to select for recombinants over non-recombinants, are stable in culture and virtually never produce chimaeric clones. BACs can incorporate up to 300kb of recombinant DNA and have all the advantages of propagation in bacterial cells (high cloning efficiency, stable in culture and easy isolation of pure DNA).

Contigs of YACs can be assembled by marker analysis and the ordering of sequence tagged sites (STSs), also fluorescent in-situ hybridisation (FISH) and the isolation and mapping of YAC end clones. Gaps in the contig can be filled in by cosmids, which can incorporate about 40kb of sequence (Collins and Hohn, 1978). The size of cosmids also makes them a useful tool for finer scale mapping. It may be necessary to subclone further if a map of still finer scale is required. Use can then be made of bacteriophage λ vectors (up to 15kb of sequence) and bacterial plasmid vectors. Cloning into the single stranded vector M13 is useful prior to sequence analysis; this approach is currently used in the large-scale 'factory-style' sequencing initiatives, such as the generation of ESTs (see section 1.8.5) and the sequencing of the human genome.

YACs are currently the vector of choice for the large scale cloning of the human genome and ultimately the production of contigs spanning every chromosome. Due to the relative difficulty of cloning human DNA into YAC vectors, much reliance has been placed on a small number of well-documented reference libraries, for example from ICI (Anand *et al*, 1990), the ICRF (Larin *et al*, 1991) and CEPH (Dausset *et al*, 1992). This has aided the process of standardisation and facilitated collaboration between groups. Much of the human genome YAC contig is already in place, although some regions are far more thoroughly mapped than others. For instance a contig of

overlapping cosmid clones spanning the entire long arm of chromosome 21 was described three years ago (Chumakov *et al*, 1992).

1.8.5 Methods for finding genes

Of the approximately 60, 000 to 100,000 genes in the human genome, only approximately 175 have been mapped, cloned and associated with a phenotype. This is not due to limitations in mapping procedures - there is now a linkage map of the whole human genome with markers spaced on average 2.9cM apart (Gyapay *et al*, 1994). Additionally many regions have been cloned into YACs and cosmids and contigs assembled. The rate limiting step in positional cloning is usually the identification of coding sequences from within regions of genomic DNA.

There are several features of genes which can be exploited in order to facilitate their detection. Firstly use can also be made of preferential cross-species conservation of gene sequences. Secondly there are recognisable sequence motifs associated with genes, for instance CpG islands, splice consensus sites and the open reading frames themselves. Thirdly features of gene expression can aid gene detection, namely the presence of mRNA and the ability to make cDNA.

Zoo Blots Gene sequences have been more highly conserved throughout evolution than inter-genic DNA, and as such there is a high probability that a sequence present in several species will be part of a gene. Use of 'zoo blots' has exploited this, and has been instrumental in the detection of a number of genes, most notably the gene for Duchenne muscular dystrophy (Monaco *et al*, 1986). A severe limitation of zoo blots however is that the probe used must be of a small size. Whole YACs are too large to use as probes, since they give an unacceptably high background. Whole cosmid inserts, a tenth of the size of YACs, have been successfully applied to zoo blots, but in general probes of only a few kilobases or less are more desirable.

CpG Islands CpG islands are regions of the genome particularly rich in the frequency of CpG dinucleotides. The frequency of CpG across a whole vertebrate genome is only approximately 20-25% of that expected (Bird, 1987). This is largely due to methylation of 60-90% of cytosines, which may be then deaminated to thymines. This mismatch is probably not recognised by the DNA repair machinery. It has been found however that regions exist at the 5' end of most genes, which have more than 10 times the density of unmethylated C+G's than the genome as a whole (Bird, 1987). These regions are termed CpG islands, or HTF (*Hpa II* tiny fragment) islands. They are between 500bp and more than 2kb long, and often span the 5'-most exons of the gene. It has been estimated that all housekeeping genes and 40% of tissue specific genes are associated with CpG islands (Larsen, 1992).

The presence of CpG islands as gene 'markers' can usefully aid gene detection. CpG islands can be detected by the presence of rarely cutting restriction enzyme sites. These enzymes have recognition sequences composed entirely or largely of C and G nucleotides, and will only cut unmethylated sequences. Thus the majority of the recognition sequences will occur within CpG islands. For instance the enzyme *Not I* has an average of 0.12 cutting sites per island, but 89% of these sites occur in an island. Similarly the enzymes *Sac II*, *BssH II* and *Eag I* have on average 1.2 sites per island, and 74% of their sites will be in an island (Lindsay and Bird, 1987). Diagnostically it can be assumed that a short region containing two or more sites for rare cutter restriction enzymes is highly likely to be a CpG island.

Rare cutter restriction enzyme digestion of genomic regions can quickly identify candidate regions for further analysis. Hanson *et al* (1991) used this principle to identify and clone clusters of rare cutter restriction sites from the HLA region of chromosome 6p. Others have developed PCR-based

strategies to isolate the genomic regions flanking CpG islands. One method, termed 'Island Rescue PCR', involves the digestion of DNA with rare cutting enzymes, then ligation of vectorettes onto the cut cohesive ends. PCR is then performed between the rare cutter site (vectorette primer) and an Alu repeat sequence (Alu consensus primer). The PCR products can then be used as probes on Northern blots or to isolate cDNA clones. Patel *et al* (1991) originally described this method as applied to somatic cell hybrids. Valdes *et al* (1994) then applied it successfully to YACs. This method obviously depends on the presence of an Alu repeat element at a suitable distance from the CpG islands and in the correct orientation.

Most recently, Cross *et al* (1994) have described a method to purify CpG islands from a whole genome using an affinity column which binds methylated DNA. The column consists of the rat methyl binding protein, MeCP2, attached to a physical support. The DNA is digested with the restriction enzyme *Mse I*, whose recognition site (TTAA) is rarely found within CpG islands. Next, fragments containing clusters of methylated CpG's are 'stripped' from the genome, by several rounds of purification on the column. The CpG island regions are then methylated by the action of the bacterial methyltransferase enzyme, which enables them to be selectively retained by the column. The CpG islands can be eluted by the addition of high-salt buffer.

Exon Trapping and Amplification The vast majority of genes in eukaryotic DNA are composed of an intron-exon structure, the introns being spliced out of the transcribed hnRNA to make a mature mRNA. The presence of splicing indicates definitively the presence of a gene. The two methods of exon trapping and exon amplification have been developed to detect genomic DNA sequences containing splice sites. Exon trapping (Duyk *et al*, 1990) involves shotgun cloning of DNA fragments into a retroviral shuttle vector which contains a splice donor site. The vector is then transfected into

mammalian cells. Pairing between vector splice donor and any genomic splice acceptor sites will result in the production of a novel mature RNA, which can be amplified by RT-PCR. This rather laborious method has now been largely superseded by exon amplification (Buckler *et al*, 1991), a slightly simpler version, which requires the presence of both a splice donor and a splice acceptor site in the genomic DNA. Exon amplification is currently viewed as one of the most efficient ways to detect genes, and has been successfully applied in the cloning of the Neurofibromatosis type 2 gene (Trofatter *et al*, 1993) and the Huntington's disease gene (The Huntington's disease collaborative research group, 1993) among others. However it is limited by the small amount of DNA which can be inserted into the (usually plasmid) vector, and not all sequences can be equally propagated. In addition, exon amplification cannot detect genes without introns, and exons with cryptic splice sites,. It is also technically demanding due to a requirement for multiple transfections of animal cells.

Other methods to find genes also use the presence of splice sites. Liu *et al* (1989) have designed degenerate hexanucleotide primers to splice donor sites to prime cDNA synthesis from hnRNA in somatic cell hybrids. The cDNA library thus constructed was then screened with total human genomic DNA to identify those clones containing a human repeat. The method used hnRNA rather than mature RNA since most human repetitive elements are located in introns rather than exons. An alternative approach (Melmer and Buchwald, 1992) is to use oligonucleotides corresponding to the splice site consensus sequences as probes for hybridisation directly onto cosmid libraries. Degenerate oligonucleotides of both 5' and 3' splice sites were used, so to increase the proportion of intron-exon boundaries detected.

Sequencing The direct sequencing of large genomic regions and subsequent detection of coding sequences by computer is rapidly becoming a feasible approach, as automated sequencing and computer algorithms to detect

genes improve. Direct sequencing seems likely to supersede many others gene finding methods in the next few years, although to date it has been largely applied to model systems. The program GRAIL is estimated to be able to detect more than 90% of exons present in a given sequence, with approximately 20% false positives.

The conventional approach involves subcloning of a region into M13, constructing a highly redundant contig, sequencing the individual clones, and performing more primer-directed sequencing to fill in the gaps before the computer algorithms can be applied. The contig assembly in these projects is a laborious task and it may not be necessary - Claverie (1994) has predicted that a 2-fold coverage of a region by random fragment sequencing *before* assembly of a contig gives a 90% probability of detecting the exons with GRAIL.

Hybridisation to Northern blots and cDNA libraries Expression dependent methods of gene finding vary considerably in their complexity and usefulness. The simplest is hybridisation of probes onto Northern blots or cDNA libraries. An obvious limitation is that the gene must be expressed, and at a sufficiently high level in the tissue from which the RNA or cDNA has been made. These methods are also generally not appropriate for large probes, meaning that coverage of a large genomic region has to be by sequential hybridisation of smaller fragments. Cosmid inserts have successfully detected transcripts on Northern blots; these represent a practical upper size limit on most occasions. The use of whole YACs as probes onto cDNA libraries has been applied with some success. Pre-annealing of the highly reiterative component of human genomic DNA to the denatured YAC is used to block high copy repeat sequences which will lead to false positives. Elvin (1990) used a 180kb YAC as a probe to detect a cDNA derived from the aldose reductase gene. The false positive rate was high, only 10% of the positive clones were also positive with an aldose

reductase cDNA probe; the aldose reductase cDNA sequence did however represent only ~1% of the probe.

Subtractive cDNA hybridisation Many subtractive cDNA hybridisation techniques exist. Jones *et al* (1991) have developed a method to isolate sequences transcribed in one hybrid cell line but not another. The second cell line contains the same complement of human DNA as the first except the region of interest. Small fragments of oligo-dT primed cDNA made from the first cell line are hybridised to biotinylated poly(A)+ RNA from the second. Any cDNA from a region contained in both cell lines will form a duplex and can be removed by binding of the biotin residue to a streptavidin matrix. After a second round of subtraction the remaining cDNAs can be used to screen a cDNA library.

Coincidence Cloning

cDNA selection by PCR The various methodologies which fall under the heading of 'coincidence cloning' involve the selection, in a clonable form, of a duplex formed between two DNA sources. In the context of gene finding, one of these DNA sources would be cDNA and the other a genomic region, such that the products isolated would be transcribed sequences mapping to that particular genomic region. The subset of techniques known as 'cDNA selection' involve hybridisation of a whole cDNA library onto immobilised genomic DNA (YACs or cosmids). One of the two resources must be pre-blocked for high-copy repeats. In the first instance, the immobilisation of the YACs was on a nylon filter, and the non-specifically bound cDNA species could be washed off before elution and PCR amplification of the product. In one study (Lovett *et al*, 1991), the DNA sources were an amplifiable foetal kidney cDNA library (pre-blocked for high copy repeats) and a 550kb YAC. The only fundamental methodological difference in the second study (Parimoo *et al*, 1991) was that the immobilised genomic DNA, not the cDNA, was quenched of available high copy repeats.

The efficiency of coincidence cloning experiments can be expressed in terms of an 'enrichment factor', which is the frequency of a particular DNA sequence in the product relative to its frequency in the starting material. In these experiments the enrichment factor was ~1000-fold, which is insufficient to recover rarer transcripts. The focus of cDNA selection experiments therefore changed to solution rather than filter hybridisation, which produced higher enrichment factors and in which the duplex-forming reaction was more easily controlled. The denatured genomic DNA is attached to streptavidin-coated magnetic beads by way of a biotin moiety added during amplification.

The pre-blocked, amplified, denatured cDNA is added before (Korn, 1992) or after (Morgan, 1992) the genomic DNA fragments have been immobilised. cDNAs forming stable duplexes with sequences present in the genomic DNA will be retained during the washing steps and may then be eluted, amplified and cloned. Further rounds of selection can be applied to increase the enrichment still further.

End Ligation Coincident Sequence Cloning (EL-CSC) EL-CSC is an extension of standard cDNA selection techniques, which can result in even higher enrichment factors (Brookes, 1994 and Brookes *et al*, 1994). This is due to the incorporation of an extra step before elution of the product cDNA, the high stringency ligation of 'capture' oligonucleotides to the ends of the cDNA while it is present in a duplex with the genomic DNA. The capture oligonucleotides are complementary to linkers present at the ends of the genomic DNA, and thus the ligation reaction can only take place if the ends of the cDNA and genomic DNA are perfectly matched both in length and end-sequence. Using EL-CSC six genes were isolated from pairing a human foetal brain cDNA library with 260kb of pooled cosmids (Brookes at al, 1994). Enrichment factors of greater than 10^6 fold can be achieved with EL-CSC, which is at least 10-fold greater than those achievable with most cDNA selection protocols. The spectrum of products generated by the two methods

is however slightly different, so they should be viewed as complements rather than alternatives.

Physical cDNA selection One of the great assets of coincidence cloning technology relative to other gene finding methods is its ability to cope with a wide variety of input genomic DNAs, both cloned and uncloned. It can even be extended as far as the use of whole chromosomes, for instance in the technique termed 'Preparative In-Situ Hybridisation' (Prep-ISH; Hozier *et al*, 1994). An amplified cDNA library is used as a fluorescent in-situ hybridisation (FISH) probe onto an array of chromosomes fixed on a slide. After removal of non-specifically bound fragments by washing, the region of coincidence can be recovered by microdissection and microcloning. Since the human genome can be resolved into up to 1000 recognisable chromosome bands, an average cDNA library of 10,000 transcripts could in theory be reduced in complexity to only 10 transcripts mapped to a particular region.

One of the main problems of coincidence cloning is the recovery of artefactual products. High copy repeat elements can be adequately removed by pre-blocking, but this does not remove lower copy repeats such as MERs (Lovett, 1994). cDNA fishing may also legitimately recover other members of a gene family or pseudogenes. More complex input DNAs compromise its efficiency, and for this reason fewer artefactual products are recovered when cosmids rather than YACs are used as the genomic input DNA (Lovett, 1994). YACs are problematic to isolate in a pure form, and there will inevitably be contamination with yeast sequences to a certain extent. Parts of the yeast genome show sufficient conservation with human DNA that they may be genuinely recovered by coincidence; this is particularly important in the case of ribosomal RNA (rRNA) sequences which are abundant in the yeast genome. Pre-blocking with an excess of denatured rRNA can however effectively remove the problem (Lovett, 1994).

The quality of the cDNA resource is also of prime importance. It is generally found that oligo-dT primed cDNA gives better results than a cDNA library since the latter are often plagued with artefacts, and cross-homology can occur between the vector sequences of the two input DNAs. The tissue of origin must be carefully chosen to maximise the number of genes expressed there. The advantage of EL-CSC is that it can theoretically recover genes expressed at a very low level. Since coincidence cloning is heavily dependent on PCR, the drawbacks of this technique must be recognised. Amplification of the starting materials will result in a biased representation, since the amplification favours shorter fragments and certain sequences. Errors in sequence will also be introduced by the polymerase, though this can be limited by including an enzyme with proof reading activity.

Whole genome efforts

Methods for detecting genes based on their chromosomal location ('top down') are now complemented by an alternative ('bottom up') strategy based upon direct sequencing of cDNAs. The latter has been developed in the last few years, has been undertaken on a huge scale, and it is likely to be more instrumental in the establishment of a primary gene map of the human genome than the conventional positional cloning approach.

This 'bottom up' strategy is a massive programme to derive STS's for all human cDNAs, termed Expressed Sequence Tags (EST's) (Adams *et al*, 1991). ESTs are derived from the sequence of the 3' end of random cDNA clones from brain cDNA libraries. The 3' ends include the untranslated region (UTR); this was chosen because 1) 3' UTRs rarely contain introns (so the EST is of the same size in genomic DNA and cDNA), 2) the identification of a poly(A) tail and polyadenylation signal is useful for verifying true human mRNAs and 3) the 3' UTR sequence is often more gene specific than the coding sequence and can therefore be used to distinguish between members of a conserved gene family (Berry *et al*, 1995).

The cDNA clones can be sequenced rapidly using automated procedures in 'factory-style' laboratories and as a result, EST sequences are currently being added into the 'dbEST' database (Boguski *et al*, 1993) at the phenomenal rate of approximately 1500 per day (Boguski and Schuler, 1995). It is estimated that more than 50% of all human genes are now represented by at least one EST.

The rate limiting step in the establishment of a human gene map via ESTs is the mapping of their chromosomal location. Until very recently, the large-scale mapping of EST was merely to assign them to a particular chromosome by PCR amplification in a set of somatic cell hybrids each containing a single human chromosome (Polymeropoulos *et al*, 1992) or by genetically mapping polymorphic cDNAs using the CEPH reference pedigrees and database (Polymeropoulos *et al*, 1993). Now however use is being made of the increasingly extensive YAC contigs and panels of radiation hybrids - one group has reported the region-specific localisation of 318 EST-derived cDNAs, representing 308 human genes (Berry *et al*, 1995).

It is unfortunate that as more knowledge of the aetiology of human genetic diseases becomes available, then the disease-causing genes themselves become precious commodities. Controversial issues regarding the patenting of genes have arisen, as several large-scale cDNA sequencing institutions teamed with drug companies seize the opportunity of a lucrative future. Fortunately the hitherto collaborative nature of the worldwide genome project has not been in vain, since other institutions are now producing an equally large EST database with unrestricted access for other researchers.

1.9 Positional Cloning methodology used in complex disorders

1.9.1 Linkage Analysis

Parametric linkage analysis on large multiply affected families has the power to detect a locus contributing a modest susceptibility to a complex disorder. The results are expressed as a logarithm of odds (lod) score, at a certain distance (expressed as a recombination fraction, θ) from the disease locus. It is accepted that a lod score of 3 or greater demonstrates significant linkage at a particular locus (Risch, 1992); this implies an odds ratio of 1000:1 that the two loci are linked.

The affected sib-pair and affected pedigree member methods

These are non-parametric linkage methods. Non-parametric methods are more robust than parametric methods since they require no assumptions about the mode of inheritance, and thus also avoid the necessity for testing multiple models. Non-parametric methods lack some of the power of the parametric methods, yet can out-perform parametric analysis if an incorrect model is applied to the latter (Terwilliger and Ott, 1994).

Sib-pair analysis relies on the assumption that two siblings will share, on average, 50% of their alleles. However two affected siblings will be expected to share alleles tightly linked to the disease locus at a frequency greater than 50%. This can be detected in a sib-pair analysis with a sufficiently great sample size. The affected-pedigree member method examines the frequency of allele sharing between two affected members of a family, with the assumption that alleles either at or very close to a disease locus will be shared more frequently than would be expected by chance. Although it is easier to obtain larger sample sizes than for sib-pair analysis, it is strongly dependent on accurate estimates of gene frequency (Terwilliger and Ott, 1994).

1.9.2 Association Studies

In an association study, the frequency of particular alleles at a polymorphic locus in a sample of unrelated individuals is compared with the frequency of this allele in controls (Hodge, 1994). An allele which occurs more frequently with the disease phenotype than would be expected by chance can be assumed to have a genetic locus close to a gene involved in the disease. Association studies are not as powerful as parametric linkage studies, but their advantage is that pedigree data are not required. This is particularly useful in late-onset conditions where members of previous generations are likely to be deceased.

Mutational Analysis of candidate genes

A particular form of association study is to search for mutations in a candidate gene, which can then be tested for association with the disorder. The association study is of particular importance so as to test whether the mutation detected is in fact only a naturally occurring polymorphism of no pathogenic significance. For instance, Nöthen *et al* (1994) reported a null mutation in the first exon of the D4 gene, leading to a truncated protein. However it was found that this null allele occurred in about 2% of the normal population, with a similar frequency in schizophrenics and controls.

1.9.3 Testing candidate genes

Once a gene has been found which demonstrates an association with the disorder, studies must be undertaken to assess the nature of the gene in relation to the disorder. Evidence accrued from several avenues will contribute to the status of the candidate gene.

Firstly, a search can be undertaken for mutations in the gene in affected individuals. Gross deletions and large chromosome abnormalities can be detected by Southern blotting and/or cytogenetics. Other methods are required for the detection of smaller deletions and point mutations. Several of

these rely on altered electrophoretic mobility, for instance single strand conformation polymorphism (SSCP) analysis or denaturing gradient gel electrophoresis (DGGE). Others depend on cleavage at a mismatch, for instance RNase A cleavage, use of the *E.Coli* mutS protein which binds to mismatched DNA or the hydroxylamine osmium tetroxide (HOT) technique. Alternatively the mutation/polymorphism can be detected by direct sequencing in patients versus controls.

The standing of a candidate gene can also be tested by logical analysis of its expression pattern and function. Hybridisation onto Northern blots and RT-PCR analysis can reveal in which tissues the gene is expressed. RNA *in situ* is time consuming but can give more suggestion of the function of the gene by its specific spatial and temporal expression pattern. If the gene is novel then sequence analysis and database searching can reveal clues as to its function, via homologies with other known genes. If there are no homologies, the protein sequence of the gene should be analysed for sequence motifs indicative of transmembrane regions, DNA binding regions etcetera.

In a heterogeneous condition such as schizophrenia, a gene mutated in one family must be tested in other affected families. If it appears to be normal in these others, then tests should be carried out on biochemically and evolutionary related genes. One approach which may be particularly useful for complex disorders is to exploit the cross-species conservation of genes, by studying animal models. Once predisposing genes have been identified and mapped in an animal, use can be made of knowledge of syntenic regions between the animal and the human genomes to attempt to identify the human counterparts. Although the genes which predispose to disease in an animal may not be exactly the same as those underlying human disease, there are often similarities in terms of number and interactions of genes and the specific physiological process involved. Elucidation of the genes involved in an animal model is simpler than in humans due to the availability of inbred

strains in which the genetic background is known, and directed matings. An understanding of the genetic basis for the complex disorder in non-obese diabetic (NOD) mice has been achieved (Ghosh *et al*, 1993).

The mouse mutant *reeler* has recently been identified as being caused by a single gene defect (D'Arcangelo *et al*, 1995). *Reeler* is an autosomal mouse mutation causing impaired motor co-ordination, tremors and ataxia. Neurons fail to reach their correct locations in the developing brain. The mutation has been found to be in a gene, dubbed 'reelin', which resembles extracellular matrix proteins involved in cell adhesion. The study of animal model phenotypes such as this may provide clues as to the underlying genetic basis of comparable phenotypes in humans.

For a more thorough investigation of the function of a novel gene, transgenic mice could be made by inserting a mutated version of the gene (for dominant genes) or by knocking out both copies of the normal gene (for recessive genes). Any phenotype thus produced in the mouse could be analysed and correlated with phenotype seen in man. Rat transgenic animals are useful for the study of dominant negative phenotypes. Animal studies of this type have begun using the prime candidate genes in schizophrenia, neurotransmitter receptor genes. Injection into the brain of an antisense oligonucleotide directed against the D2 receptor can block D2 receptor activity by causing a steric hindrance or creating a substrate for RNase H. It was found that D2 activity was effectively blocked across the whole striatum (as measured by the drop in binding of radio-labelled spiperone) and caused a decrease in spontaneous locomotor activity in the rat (I.J. Creese, presented at the 'International Congress on Schizophrenia Research', Virginia, USA, 1995). Sasaoka and Tonegawa (1995) have produced transgenic mice which either overexpress the D4 gene or are knock-outs for the NMDA receptor R2B. The aim will be to cross the two mice to study the effect of defects in both the dopamine and glutamate systems. The work is still in the early stages, but

some hyperactivity and exaggerated startle response has been noted in the D4+ mice.

1.10 Previous Attempts to Find Genes Associated with Psychiatric Disorders

There are currently three fundamental approaches to finding genomic regions associated with psychiatric disorders (Bassett, 1991). The first is to focus on certain regions of the genome, those for which there is some *a priori* evidence for involvement with the disorder. This may include regions previously implicated in linkage studies, chromosomal regions where particular cytogenetic abnormalities are linked with psychiatric illness, or information derived from case studies in which schizophrenia is seen to segregate with another genetic disease whose locus is known. The second approach is to carry out linkage and segregation analysis across the whole genome, to find 'hot-spots' for regions likely to be linked with the disorder. Recently formed collaborative consortia between research groups world-wide, and the ever-increasing density of markers available for analysis have made this feasible. The third and final method is to study candidate genes, drawing upon knowledge from hypothetical aetiologies. Linkage and association analysis can be performed using polymorphisms detected within the genes, or a more direct search for mutations can be carried out. This approach proved extremely useful in Alzheimer's disease when mutations were found in the gene for the amyloid precursor protein (APP) (Goate *et al*, 1991).

1.10.1 Linkage in candidate regions

Chromosome 5q

The much heralded first positive linkage of a genomic locus with schizophrenia was announced in 1988 (Sherrington *et al*, 1988). The report of a partial trisomy of 5q in a schizophrenic boy and his schizophrenic maternal uncle (Bassett *et al*, 1988), generated interest in performing linkage

studies with markers from the region of the trisomy. Sherrington *et al* reported a maximum multipoint LOD score of 6.49 at markers with two RFLP markers in the region (5q11-13). These data came from analysis of 39 cases of schizophrenia from 2 British and 7 Icelandic families, with a broad phenotypic definition, dominant mode of transmission and 86% penetrance. This result has however never been replicated. In the same issue of Nature as the original paper was a report excluding linkage at 5q11-13 in a large Swedish kindred (Kennedy *et al*, 1988). Since then many other linkage studies have been carried out, all with the same negative conclusions (for example Detera-Wadleigh *et al*, 1989, St.Clair *et al*, 1989, Aschauer *et al*, 1990, Crowe *et al*, 1991, McGuffin *et al*, 1990 and Macciardi *et al*, 1992). The original report of linkage might therefore have been due to chance or to inappropriate methodology or statistics. Alternatively it is possible that the families studied segregated a rare form of the disorder. The latter is plausible for the Icelandic families, since they were geographically isolated, but seems unlikely due to the inclusion in the positive findings of the two British families.

One patient has been reported in which there is a co-occurrence of schizophrenia and Treacher Collins syndrome, which is a disorder of cranio-facial development. Two siblings of the patient also had a psychiatric diagnosis. Treacher Collins syndrome has been mapped to 5q31-34. A cytogenetic survey of the patient was performed to detect any gross chromosomal deletions of the Treacher Collins region plus the putative schizophrenia locus, but no abnormalities were detected (Nimgaonkar *et al*, 1993a).

Chromosome 11q

There have been three reports of chromosome translocations involving chromosome 11q associated with psychiatric illness. The first, a balanced translocation between chromosomes 1 and 11 (St.Clair *et al*, 1990) forms the basis for this thesis and therefore is described in detail later (section 1.12.2).

A second translocation t(11;9)(q22.3;p22) was found to co-segregate with affective disorder in at least 5 members of an extended pedigree (Smith *et al*, 1989). The chromosome 11 breakpoint was estimated to be within 210kb of the dopamine D2 receptor. Thirdly, a balanced translocation between 11q25 and 6q14.2 was reported to co-segregate with psychotic illness in a three generation family (Holland and Gosden, 1990).

Linkage analysis with other markers on chromosome 11 has however not yielded any significant lod scores. D11S35 gave a lod of 1.49 in a Japanese study (Nanko *et al*, 1992) and 1.8 (conservatively corrected score) in another study of 23 multiply affected schizophrenics (Gill *et al*, 1993).

The sex chromosomes

Evidence for the involvement of the sex chromosome in schizophrenia has arisen from an apparent excess of sex chromosome abnormalities seen in schizophrenics. Sex chromosome aneuploidies such as Turner's syndrome (XO), Klinefelter's syndrome (XXY), triple X, XYY and mosaicism for X chromosome aneuploidies have all been observed among female schizophrenics (Barr *et al*, 1994a) However a strict X linked mode of inheritance can be excluded due to the presence of male to male transmission in some families. Parametric linkage analysis and affected sib-pair analysis across the entire X and Y chromosomes yielded a positive but statistically insignificant maximum lod score of 1.83 at the marker DXS7 (De Lisi, 1994). Crow *et al*, (1993) have reported an excess of shared alleles at the androgen receptor gene locus between male sib pairs. This locus is a candidate gene for schizophrenia due to the observed sex differences in the age of onset and course of the disorder.

The pseudoautosomal region of the X and Y chromosomes has been particularly implicated. This is a region at the tip of the X and Y chromosomes which contains regions of sequence similarity between the

two, and undergoes a single obligatory recombination during meiosis. As a result, the genes at the telomere segregate independently of sex, and there is an increasing gradient of sex linkage until the boundary of the pseudoautosomal region is reached (Ishida *et al*, 1993). An involvement of the pseudoautosomal region in schizophrenia was first implicated when it was noted that there tended to be a higher concordance for schizophrenia in same-sex siblings than in opposite-sex pairs. This can be explained under a model of paternal transmission of a pseudoautosomal allele, which would be passed on the X chromosome to daughters or the Y chromosome to sons. The model also implies that the excess of concordance should be in those families where the disorder is paternally derived. Some studies have confirmed the excess of same-sex pairs (Gorwood *et al*, 1992 and Asherson *et al*, 1992), whereas others have refuted it (Ishida *et al*, 1993).

Despite this observation and many parametric and non-parametric linkage and association studies, there have been no consistent reports of linkage to markers in the pseudoautosomal region. Collinge *et al* (1989) reported a positive linkage result with the marker DXYS14, but others have failed to replicate this in other families (d'Amato *et al*, 1992 and Parfitt *et al*, 1991). Asherson *et al* (1992) found an excess of shared alleles at DXYS14 in schizophrenics but excluded this marker by linkage. An analysis of the entire pseudoautosomal region in a Swedish kindred produced no significant linkage (Barr *et al*, 1994a), although Crow *et al* (1994) reported a lod score of 2.44 at the gene MIC2, which is located within the pseudoautosomal region.

<u>The human lymphocyte antigen (HLA) region, 6p21.3</u>

The HLA region was a candidate for linkage analysis since many immunological abnormalities have been associated with schizophrenia. Changes in T cell dependent reactions have been noted, along with an increase in numbers of CD3+ and CD4+ cells. Association studies have been

performed with many HLA antigens. In a sample of 100 Belgian schizophrenics a significant negative association was found between schizophrenia and the DPB1*0101 allele (relative risk of 0.27) and a positive association was revealed with CD4*A4 (relative risk 1.79) (Zamani *et al*, 1994). A slightly decreased chance of paranoid schizophrenia has been found in those expressing the HLA-A1 allele, but an increased chance with HLA-A9, A28, B18, B27 and Cw24. Risk of hebephrenic schizophrenia is slightly associated with the A1 and B15 alleles. (Douglass *et al*, 1993)

Chromosome 2

The report of a balanced translocation t(2;18)(q21;q23) segregating with schizophrenia in a family (Genest *et al*, 1976) prompted a linkage study using markers located at 2q21. No evidence for linkage was found, and 50cM around 2q21 could be excluded (Aschauer *et al*, 1993).

Chromosome 9

A G-banding survey of schizophrenics versus controls (Nanko *et al*, 1993a), revealed an increased frequency of a pericentric inversion of chromosome 9 relative to controls (Serra *et al*, 1990). This was in concordance with an increased prevalence of the same chromosome anomaly found in male patients with paranoid psychosis (Axelsson and Wahlstrom, 1984). However a linkage study on three multiply affected families using markers from the region of the inversion (D9S55, D9S15 and D9S202) resulted in no significant lod scores (Nanko *et al*, 1994a).

1.10.2 Whole genome scans

A whole genome scan of patients from 39 systematically ascertained multiplex families (Pulver *et al*, 1994a) revealed a maximum lod score of 1.54 at chromosome 22q12-q13.1, a region which contained several candidate genes. This slight positive finding was then followed up by a collaboration of 4 groups using families from Ireland, USA, France, England, Wales and

Japan (Pulver *et al*, Diehl *et al*, Gill *et al* and Laurent *et al*, 1994b). This follow up study found no evidence for linkage in this region of chromosome 22q.

Other chromosome 22 studies have revealed no consistent results - a) The marker D22S55 (22q) gave a slightly positive lod score in a whole genome scan with 329 polymorphic loci, which remained positive when the affected pedigree member method was employed (Coon *et al*, 1994a). b) A linkage study of 9 families with 10 chromosome 22 markers gave a maximum lod score of 2.09 at D22S276. Non-parametric tests also gave suggestive but not conclusive evidence for linkage with this marker (Coon *et al*, 1994b). c) Polymeropoulos (1994) scanned 105 families with 10 chromosome 22 markers and found no evidence for linkage. d) Linkage analysis with markers D22S274 and D22S283 (22q12-13) on twenty three UK and Icelandic families yielded no significant lod scores (Kalsi *et al*, 1995).

Linkage analysis with an autosomal dominant mode of inheritance, using 180 polymorphic markers in a Swedish kindred resulted in no positive findings, but 330cM of the genome were able to be excluded from involvement in this family (Barr *et al*, 1994b).

The only significant lod score which has been produced by a whole genome scan thus far is that of Wang *et al* (1995). Linkage analysis of markers covering 10% of the genome was performed on 186 Irish families, which resulted in a maximum lod score of 3.2 at D6S260 which maps to 6p23. Use of the affected pedigree member method also gave significant results at this locus. The positive results did not include the HLA region.

1.10.3 Analysis of candidate genes

Dopamine Receptors

The genomic loci of the five dopamine receptors are prime candidates for linkage to schizophrenia due to the proposed abnormalities in the functioning of the dopamine system in the disorder. D3 and D4 are of particular significance; the former due to its restricted limbic expression pattern and the latter due to its high affinity for clozapine, the most effective antipsychotic drug. The cloning and molecular analysis of the receptor genes yielded polymorphisms which could be used for linkage analysis. For instance the D2 gene contains a polymorphic base pair which results in an alternative amino acid (serine or cysteine) at position 311 (Itokawa *et al*, 1993). D3 contains a *Bal I* RFLP (Lannfelt *et al*, 1992), and D4 has a variable number of copies of a 48bp tandem repeat in the putative third cytoplasmic loop (Van Tol *et al*, 1992). More than twenty alleles of D4 may exist as a result of this repeat, which may be of significance in schizophrenia since different alleles exhibit different binding affinities for clozapine and spiperone.

Despite many attempts, no consistent linkages or associations have been reported at any dopamine receptor locus. (D1: Campion *et al*, 1994. D2: Campion *et al*, 1994, Nanko *et al*, 1994c. D3: Yang *et al*, 1993, Nanko *et al*, 1993b, Nimgaonkar *et al*, 1993b, Nanko *et al*, 1994b, Shaikh *et al*, 1993 (bipolar affective disorder). D4: Sommer *et al*, 1993, Nöthen *et al*, 1994, Barr *et al*, 1993, Shaikh *et al*, 1994, Macciardi *et al*, 1994, Seeman *et al*, 1994b, Lim *et al*, 1994 (bipolar affective disorder)). One finding, replicated once, was an excess of homozygosity for D3 alleles in schizophrenics (Crocq *et al*, 1992). This was replicated to a lesser extent by Mant *et al* (1994), although other many other studies have found no such association (Nöthen *et al*, 1993).

In addition to linkage studies, the status of the dopamine receptor genes as candidates for schizophrenia has been examined by sequence analysis.

Sarkar *et al* (1991) sequenced 7 regions of functional significance (including the coding sequence and splice junctions) of the D2 gene in 14 schizophrenics and 4 normal controls. No associations were found.

Linkage has been carried out for markers at the dopamine transporter gene, mapped to 11p15.3, whose activity is the termination of dopamine function by reuptake into the pre-synaptic terminal. No positive lod score was found (Persico *et al*, 1995).

Other neurotransmitter receptors

Coon *et al* (1994c) examined the coding region of the $GABA_A\beta1$ subunit gene for sequence variants. SSCP analysis revealed a C to G change which would result in glutamate rather than histidine at position 396, a highly conserved portion of the gene. This variant was found in 2 out of 3 affected siblings in one family, but not in a further 155 unrelated schizophrenics; it was also found in 1.1% of controls.

Hallmayer *et al* (1992) performed linkage analysis with markers spanning 19cM around the 5HT2 gene on chromosome 13. Linkage at this locus was excluded in a Swedish kindred.

Amyloid Precursor Protein (APP)

Mutations in the gene APP has recently been discovered in early-onset Alzheimer's disease patients. Since 30-40% of Alzheimer's patients display psychotic symptoms, APP is also a candidate gene for schizophrenia. In 1992 Jones *et al* reported a single case in which a patient with schizophrenia had a point mutation in the APP gene, a C to T transversion that would result in a valine rather than alanine at position 713. This variant was however not found in a further 100 schizophrenics or 100 non-demented controls (Jones *et al*, 1992), 39 Japanese schizophrenics (Fukuda *et al*, 1993), or 86 schizophrenics (Coon *et al*, 1993a) and as such is most likely to be a rare

non-pathogenic variant.

Linkage analysis was carried out on 24 multiply affected families with a marker at the APP locus, and SSCP mutation analysis on one member from each family. Neither approach resulted in a positive result (Asherson *et al*, 1993). Linkage analysis also gave negative results with markers spaced at 20cM intervals along chromosome 21q.

No association analyses have yet been reported with other genes involved in Alzheimer's disease (for instance ApoE4 and S182) although their status as candidate genes in schizophrenia has been mentioned (Goodman, 1994).

Tyrosinase

There have been several reports of families in which schizophrenia is associated with tyrosinase negative oculocutaneous albinism (OCA). Baron (1976) described a 22-member family of Yemenite-Jewish origin, among which 5 individuals were diagnosed with OCA and schizophreniform psychosis. Another report described a link between OCA and schizophrenia in 2 out of 3 siblings of an Irish family (Clarke and Buckley, 1989). It is possible that this association reflects the chromosomal location of a gene predisposing to schizophrenia, that is near tyrosinase on chromosome 11q. Alternatively a metabolic abnormality or the disordered neuroanatomy associated with albinism may predispose to schizophrenia.

Porphobilinogen Deaminase (PBGD)

The gene PBGD is involved in acute intermittent porphyria, which makes it a candidate gene for schizophrenia since individuals suffering from porphyria sometimes display signs of mental disturbance reminiscent of schizophrenia. An association study (Sanders *et al*, 1991) using a dichotomous PBGD RFLP revealed a relative risk for schizophrenia of more than 2 in individuals possessing the A2 allele, and an earlier age of onset in A2 homozygotes.

Later Owen *et al* (1992) attempted to replicate this study but found no positive associations.

Genes involved in Usher's syndrome

A sibship of 8 has been reported in which there were 3 cases of Usher's syndrome and 6 cases of mental retardation (Sharp *et al*, 1994). In addition, many other members of the family have psychiatric disorders. The primary symptom of Usher's syndrome is deafness, so it is of interest that schizophrenia is associated with abnormalities in auditory information processing. A gene for Usher's syndrome Type1B has recently been discovered to be myosin Type VII, mapping to chromosome 11q (Weil *et al*, 1995).

Phenylalanine hydroxylase (PAH)

Lack of PAH is responsible for phenylketonuria (PKU). The gene is a candidate gene in schizophrenia for the following reasons: a) Patients with PKU and their first degree relatives often display behavioural abnormalities including psychosis. b) In the absence of PAH, increased plasma and urine levels of phenylethylamine (PEA) are observed. The same has been observed in schizophrenics. c) PEA potentiates dopamine and norepinephrine in the same way as amphetamines (reviewed in O'Reilly and Davis, 1993).

An association study between mutations of PAH and schizophrenia however proved negative (Sobell *et al*, 1993).

Dystrophin

A Brazilian family has been described in which schizophrenia cosegregates with Becker muscular dystrophy (Zatz *et al*, 1991). Southern blot analysis however revealed no detectable large exonic deletions (Lindor *et al*, 1994). In addition, 2 out of a series of 35 patients with Becker muscular dystrophy

were diagnosed also with schizophrenia (Beggs *et al*, 1991).

Huntingtin

The symptoms of schizophrenia are seen in 5-10% of Huntington's disease patients. One schizophrenic (non-HD) individual has been reported whose number of trinucleotide repeats falls outside the normal range and within that for Huntington's disease (St.Clair, 1994). In another study the schizophrenic probands all had repeat lengths at the HD locus within the normal range (Rubinsztein *et al*, 1994).

1.10.4 Linkage studies in bipolar affective disorder

Attempts to locate genes implicated in bipolar affective disorder has mirrored the efforts in schizophrenia. A significant linkage was first detected in an Old Order Amish population, at the Harvey ras and insulin loci on 11p15 (Egeland *et al*, 1987). This result was however not replicated in Icelandic (Hodgkinson *et al*, 1987) or North American Irish families (Detera-Wadleigh *et al*, 1987), and later the finding was withdrawn by the original authors. Completion of typings, further linkage analysis upon extensions of the pedigree and onset of the illness in two family members negated the result (Kelsoe *et al*, 1989).

Despite exclusion of linkage to 11p15, a positive association was reported for the tyrosine hydroxylase (TH) gene in a French population (Leboyer *et al*, 1990). Interestingly enough, TH maps to 11p15. TH is a good candidate gene in psychiatric illness since it is a rate limiting enzyme in catecholamine synthesis. This finding has however not been replicated (Lim *et al*, 1993).

The other major locus implicated in bipolar affective disorder is Xq28, containing the loci for deutan colour blindness and glucose-6-phosphate dehydrogenase (G6PD). There have been repeated reports of linkage to these two classical markers, one study generating a lod score of nearly 8 for

a highly penetrant dominant gene (Baron *et al*, 1977). More recently, linkage has been reported at the blood clotting Factor IX locus (Mendlewicz *et al*, 1987). However, this gene lies on Xq27, a sufficiently great distance from the other positive loci that it cannot be seen as a replication of the earlier findings.

Whole genome scans have also been attempted for bipolar disorder. In one study (Coon *et al*, 1993b), eight multiplex families were examined for linkage with 328 markers across the genome. Four markers gave a lod score of more than 1 (D5S39, D5S43, D5S62 and D11S85), but only D5S62 remained positive upon analysis by the affected pedigree member method. This marker maps to distal 5q, in a region containing several neurotransmitter receptors (D1, adrenergic receptors α1 and β2, GABA receptor α1 and a glutamate receptor) which are good candidate genes in schizophrenia (Coon *et al*, 1993).

1.11 Problems with Linkage Studies in Complex Disorders

It can be seen from the wealth of negative or at least unconvincing results from linkage and association studies to date, that there has been no single locus identified unequivocally to be involved in the aetiology of schizophrenia or other psychoses.

For single locus genetic traits, conventional linkage studies are extremely adept at detecting genuine linkage between the disease locus and a marker. To date, positive linkage scores have been generated in more than 400 cases. For the following reasons however, basic linkage analysis is not so effective for complex traits such as schizophrenia.

1) Linkage analysis requires assumption of a single gene model, whereas a polygenic system with modifications from the environment is more likely (Baron *et al*, 1990). A variety of linkage studies must be carried out under

different models of inheritance, for instance dominant, recessive or mixed, which may not intuitively fit the pattern of segregation seen in the family.

2) Gene frequency and penetrance must be specified. Again, a range must be used. Although misspecification of genetic parameters will usually not prevent the detection of linkage, it will result in an erroneous recombination frequency between the marker and the disease locus (Cloninger, 1994).

3) Diagnostic boundaries are critical and must be specified in advance. Nothing is known concerning the aetiological homogeneity of the schizophrenia spectrum disorders and so it is unclear which disorders should be grouped together for linkage analysis (Baron *et al*, 1990). False positive inclusions will increase the estimated genetic distance between the marker and the disease until it becomes so great that a genuine linkage is not detected. General guidelines for inclusion have been formulated from the results from family studies (Levinson and Mowry, 1991). Probands must have a diagnosis of strict schizophrenia, with a course of at least two years. Relatives are included as affected if they have been diagnosed with schizoaffective disorder (but not if more than two relatives have schizoaffective disorder), schizotypal and paranoid personality disorder, schizophreniform illness, atypical psychosis and delusional disorder. Families are not excluded if they contain a case of bipolar affective disorder or depressive psychosis or if the inheritance of schizophrenia appears to be bi-lineal (Levinson and Mowry, 1991).

In some cases the power to detect linkage in families can be increased by including those affected by the neurocognitive traits previously discussed as 'markers' for schizophrenia (section 1.4.10) (Keefe *et al*, 1991). These may include eye-tracking dysfunction, attentional deficit and abnormalities in the P300 and P50. In future studies, some use may be made of the

neuroanatomical changes seen in some schizophrenics, and the decrease in cortical glucose utilisation.

Normally a range of linkage analyses are required using broad to narrow definitions of the schizophrenia spectrum.

4) The aetiology of schizophrenia may be heterogeneous (Baron *et al*, 1990). The enormous variety in symptoms and course of illness plus the many proposed aetiological explanations does indeed suggest that heterogeneity may exist. In addition, phenocopies of the disorder may be caused by substance abuse and organic brain disorders such as temporal lobe epilepsy, Huntingdon's disease and porphyria. Approximately 25-30% of carefully diagnosed schizophrenics show evidence of organic impairment with CT scans (Faraone and Tsuang, 1985).

5) Assortative mating occurs more frequently between psychiatric patients and as such may cause aetiologically distinct forms of the disorder to segregate in one family (Rieder and Gershon, 1978).

6) The number of affected individuals required to detect linkage is great. An estimated 200 sib-pairs are needed in order to achieve 80% power to detect linkage with a LOD of 3 at θ=0.1 with a fully informative marker (Kruglyak and Lander, 1995). One large, multiply affected pedigree alone may contain sufficient power to detect linkage, but the family may be segregating a rare form of the disease, and the linkage is critically dependent on the diagnosis of key family members (Cloninger, 1994). Large, multiply affected families are also hard to find due to incomplete penetrance, decreased reproductive fitness and high geographical mobility. Families containing more than one schizophrenic member are often unstable and become disrupted (St.Clair *et al*, 1994). Reliance may then be placed on sketchy and sometimes unreliable case notes and information from relatives concerning the diagnosis of absent

or deceased family members. Small families may have fewer complicating factors such as assortative mating and phenocopies, but larger numbers are required.

1.12 Current Strategies To Find Genes Involved In Psychiatric Disorders

It has become evident from the plethora of negative linkage results that classical methods for detecting linkage in single gene disorders may well be inappropriate for use in complex disorders such as schizophrenia. There are two fundamental ways to overcome this problem. Firstly the methodology used for linkage studies of unrelated individuals can be improved. Second is the identification of simple systems in which the disorder can be clearly demonstrated to be linked with a physical marker. The best example of the latter has been the identification of a large single family in which mental illness segregates with an unambiguous cytogenetic abnormality, a balanced translocation between chromosomes 1 and 11.

1.12.1 Methodology of linkage studies

There must be large samples of pedigrees, systematically ascertained with standardised procedures to aid collaboration and replication (Cloninger, 1994). A range of standardised inclusion criteria should be implemented, and the models decided upon prior to analysis (Baron *et al*, 1990). When possible, attempts should be made to separate subtypes of schizophrenia into aetiologically distinct varieties (for instance subgroupings according to age of onset and mode of inheritance produced positive linkage for Alzheimer's disease) (Cloninger, 1994).

Use must be made of the more sophisticated linkage methods now available, which can deal with a range of affectedness rather than a simple dichotomy of affected versus non-affected. Complex methods of segregation analysis can now test for single locus and polygenic components, while taking into

account reduced penetrance, variable age of onset and phenocopies. As has been learned from linkage analysis for Insulin-dependent Diabetes Mellitus (IDDM), it may be necessary to combine segregation and linkage approaches in order to detect a genuine association (Julier *et al*, 1991).

Replication by independent researchers is critical. Standardised procedures will facilitate replication. Confirmation of linkage in the light of more markers, updated diagnoses and extension of the pedigrees will increase the validity of any linkages detected.

1.12.2 A family in which mental illness is strongly associated with a balanced translocation, t(1;11)(q42.2;q21).

A large Scottish pedigree has been reported (St.Clair *et al*, 1990), in which a balanced translocation between chromosomes 1q and 11q has been found to segregate with psychiatric illness. The proband was first ascertained in 1968 during a cytogenetic survey of boys admitted to Scottish borstals. Information regarding other members of the family was accrued through the MRC Human Genetics Registry which contains information and annual clinical follow-up data on 282 pedigrees with familial autosomal abnormalities. Seventy-seven family members were then included in a detailed analysis, of whom 58 were still living. Diagnoses of all those with major mental illness were evaluated by a psychiatrist blind to karyotype status, and 30 of the 58 living family members were interviewed and rated according to the structured criteria of the SADS-L. Information regarding the other 28 living family members was ascertained by study of case notes and discussion with general practitioners and relatives.

It was found that 23 members of the family met Research Diagnostic Criteria for mental and or behavioural disorders. There were 3 cases of schizophrenia, 2 of schizoaffective disorder, 6 of recurrent unipolar major depression, 3 of generalised anxiety disorder, 1 of minor depressive disorder

and 3 of alcoholism. Among these, 2 had committed suicide and 3 attempted suicide. A further 5 members of the family had or were receiving treatment for adolescent psychiatric disorders (including severe conduct disorder, chronic generalised anxiety, learning difficulties and educational disturbance, bizarre behaviour and mixed adolescent conduct and emotional disorders.

Karyotyping of the 77 family members revealed that 34 carried the t(1;11) rearrangement. 16 of these 34 (49%) had a psychiatric diagnosis. In contrast, only 5 of the 43 (12%) members without the translocation had a psychiatric diagnosis, and none of these were major mental illness (1 was generalised anxiety, 1 minor depressive disorder and 3 alcoholism). It is interesting to note that the only psychiatric diagnoses present in family members not carrying the translocation are ones not normally included in the schizophrenia spectrum of presumed aetiological homogeneity.

Lod scores against chance linkage of the translocation with the mental illness were calculated. A range of gene frequencies and penetrances were implemented. Under a narrow model including solely schizophrenia and schizoaffective disorder the lod score was 2.19, but it rose to 3.33 with the inclusion of major depressive disorder. Using a broad definition of affected status, incorporating the above diagnoses plus adolescent conduct and emotional disorder, the maximum lod score of 4.34 was achieved. When generalised anxiety, minor depression and alcoholism were included in the affected phenotype, the lod scores fell to below significance.

This provides strong evidence that there is a dominant locus predisposing to major mental illness in this family near one or both of the translocation breakpoints. There has been prior reason to assume the existence of a predisposition locus on chromosome 11q. Firstly there are two other 11q translocation breakpoints associated with psychiatric illness. There are also several candidate genes located on 11q, such as the dopamine D2 receptor

gene and a neural cell adhesion molecule (NCAM) gene (St.Clair *et al*, 1994). Associations have been reported between psychiatric illness and albinism and also porphyria; the genes responsible for these illnesses (tyrosinase and PBGD respectively) are located on 11q. Finally Usher's syndrome has been mapped to near the breakpoint and a family has been described in which schizophrenia co-segregates with Usher's syndrome (St.Clair *et al*, 1994).

1.13 Experimental Background to the Project

Since there were many reasons to suggest that a locus predisposing to schizophrenia lies on chromosome 11q, it was decided that this side of the translocation breakpoint should be cloned as a priority. Lymphoblastoid cell lines were set up from translocation carriers and non-carriers in the family. Genetic linkage analysis was carried out with available polymorphic markers in the region, using markers from both chromosomes 1 and 11. On chromosome 11 no recombinations were found with tyrosinase (TYR) and the anonymous markers D11S388, and FISH analysis revealed that they were the closest known markers flanking the breakpoint (Fletcher *et al*, 1993). The distance between TYR and D11S388 has been estimated at 7.2cM (NIH/CEPH collaborative mapping group, 1992), a distance too great to begin positional cloning without the aid of more markers. To this end, the translocation-derived chromosomes were segregated into separate mouse-human somatic cell hybrids (MIS7 and MIS39) using cell surface marker selection strategies (Fletcher *et al*, 1993). Analysis of markers on these hybrids confirmed the finding that TYR was the closest known marker above the breakpoint, and D11S388 the closest below. Further mapping resources were generated by X-irradiation of the chromosome 11-only hybrid J1, to create stable chromosome fragment cell lines. The human DNA content of these four WJX hybrids was assessed by PCR marker analysis and FISH with Cot 1 human DNA (Fletcher *et al*, 1993).

To generate markers in the interval between TYR and D11S388, the technique of microdissection and microcloning was used. This technique exploits the fact that the chromosomal abnormality can be visualised under phase-contrast microscopy; the derived chromosome 1 contained a discontinuity on the long arm of 1q. Two microdissection experiments were undertaken (Muir *et al*, 1995). In the first, 160 fragments derived from the translocation breakpoint region were microdissected, digested with EcoR1 and cloned into lambda vectors. This led to over 400 recombinant clones with an average insert size of 0.3kb. The second experiment yielded over 600 clones with a larger insert size, average 1.5kb. PCR amplification of these clones with vector primers gave a total library length of more than 1Mb (ignoring redundancy). The clones from both libraries were screened with a $(GT)_{25}$ dinucleotide repeat probe, and five positives were found in the second library (Muir *et al*, 1995).

A set of 86 repeat-free microclones from the second library were then used as hybridisation probes onto a Southern blot containing the WJX chromosome 11 fragment hybrids and MIS7.4 (the hybrid containing the derived chromosome 1). Forty-nine of these were assigned to chromosome 11 and 37 to chromosome 1, approximately half each, which confirms the validity of the microdissection approach. A pool of all the microdissection clones from each library was labelled as a FISH probe and 'painted' back onto metaphase spreads from a translocation patient in the family. Signal was obtained on the normal chromosome 1 (band 1q42.2) and 11(band 11q21) plus the derived chromosome 11 as expected (Muir *et al*, 1995).

The 49 clones from the second experiment mapping to chromosome 11 were assigned to regions according to results of mapping them to the somatic cell hybrid panel. Clones mapped to all the regions defined, with the exceptions of the regions defined by NCAM and DRD2 (Evans *et al*, 1995). They also defined 2 new intervals on chromosome 11. Eleven of the microdissection

clones however map to the region immediately distal to the breakpoint, as defined by the marker D11S873, a recently defined marker that lies centromeric to D11S388 (Litt *et al*, 1993). The distance between TYR and D11S873 can be estimated by subtraction to be ~5.6cM (NIH/CEPH Collaborative Mapping Group, 1992), and hence the eleven novel microdissection clone markers within this region should provide a sufficiently dense framework to construct a YAC contig across the region.

Five repeat free microclones from the eleven (MD220, MD283, MD176, MD122 and MD471) were used to screen gridded libraries of YAC recombinants (Evans *et al*, 1995). Two libraries were screened, the ICI YAC library (Anand *et al*, 1990) (courtesy of the UK HGMP Resource Centre) and the ICRF Reference library (Larin *et al*, 1991) (courtesy of Dr. Hans Lehrach). Twenty three independent YACs were isolated in this way. These were screened for positivity with the eleven microclones and thus a simple first generation contig could be assembled, with just a single gap This however contained no information about the orientation of the contig or the extent of overlap between YACs (Evans *et al*, 1995).

1.14 Aims of this Thesis

1) To isolate end clones from the YACs efficiently, developing and using splinkerette PCR (an alternative to vectorette PCR). This to include a demonstration of the superior function of splinkerettes compared with vectorettes.

2) To map the end clones isolated back onto the YACs, thus assessing the extent of overlap between the YACs and enable the construction of a detailed contig of the region.

3) To map some of the YACs in the contig by pulse field gel electrophoresis to aid in the production of a large scale restriction map of the region and a search for CpG islands.

4) To undertake a search for genes in the region nearest the breakpoint, using hybrid-fishing coincident sequence cloning (HF-CSC).

5) To analyse the results of the HF-CSC experiment in comparison with a parallel end-ligation coincident sequence cloning (EL-CSC) experiment.

6) To map the products isolated onto the genomic region of interest.

7) To prove the genic nature of products identified and to begin to assess their rôle as candidates for a gene predisposing to psychiatric disease.

CHAPTER 2

MATERIALS AND METHODS

All chemical were supplied by BDH unless otherwise stated.

All solutions and methods were taken from or adapted with minor modifications from Sambrook *et al* (1989) unless otherwise stated.

2.1 Bacterial Cell Culture and Plasmid DNA Preparation

2.1.1 Media and solutions

All media were sterilised by autoclaving.

Terrific broth

In 900ml water dissolve 12g tryptone (Difco), 24g yeast extract (Difco), 4g glycerol. Autoclave then add 100ml autoclaved phosphate solution (0.1M KH_2PO_4 & 0.72M K_2HPO_4).

L-agar

In 1 litre water dissolve 10g tryptone, 5g yeast extract, 10g NaCl, 2.46g $MgSO_4$, 15g agar (Oxoid Ltd).

Ampicillin (Sigma)

Make stock at 50mg/ml in dH_20, filter sterilise. Store at -20°C.

Use at final concentration 50µg/ml.

Tetracycline (Sigma)

Make stock at 5mg/ml in 100% ethanol. Store at -20°C.

Use at final concentration 50µg/ml.

X-Gal (Sigma)

Make stock at 20mg/ml in DMF (Sigma). Store protected from light at -20°C.
Use at final concentration of 40µg/ml.

IPTG (Sigma)

Make stock at 100mM. Store protected from light at -20°C.
Use at final concentration of 0.5mM.

Hogness Solution

3.6mM K_2HPO_4, 1.3mM KH_2PO_4, 2mM $Na_3Ci.2H_20$, 1mM $MgSO_4.7H_20$, 4.4% glycerol.
Make stock at 10x concentration. Filter sterilise.

GTE

50mM glucose, 10mM EDTA, 25mM Tris.HCl (pH 8)
Use autoclaved Tris and EDTA. Sterile filter the glucose.

2.1.2 Growing bacterial cells on agar plates

The desired volume (up to ~200µl) of bacterial cells is pipetted onto the surface of the L-agar, and is then spread by a sterile bent glass rod until the liquid has thoroughly soaked into the agar. (If more cells than are contained in 200µl are required, the cells may be concentrated by 1 min centrifugation in a microfuge followed by resuspension of the pellet in a smaller volume.) The plates are then inverted and incubated for 12-16 hours at 37°C. Cells were viable from plates stored at 4°C for several weeks.

2.1.3 Frozen stocks of bacterial colonies

The selected colony was restreaked on a fresh L-agar plus ampicillin plate, then a single colony from the new plate was used to inoculate 5ml of Terrific broth plus ampicillin. After incubation at 37°C for approximately 16 hrs, a 1/10 volume of 10x Hogness solution was added to 1ml of the culture in a

standard 1.5ml eppendorf tube (Treff). This was then frozen and stored at -70°C.

To reactivate cells from frozen stocks, a plastic loop was passed across the surface of the frozen culture and sufficient cells would be picked up to spread on an L-agar plate and grow at 37°C.

2.1.4 Use of colony picker

A colony picking robot (Hybaid) was used according to the manufacturer's instructions to spot the CSC product library colonies onto filters and in the preparation of their frozen stocks.

2.1.5 Preparation of plasmid DNA (small scale)

The method used was a modification of the alkaline lysis method of Jones and Schofield (1990).

The desired colony on an agar plate was used to inoculate 6ml of Terrific broth plus ampicillin and was grown at 37°C overnight with continuous shaking at 250rpm. The culture was then centrifuged for 1min in a standard 1.5ml eppendorf tube, adding 1.5ml of the culture at a time and retaining the pellet. The pellet was then resuspended in 200µl GTE (50mM glucose, 10mM EDTA, 25mM Tris.HCl (pH 8) and left on ice for 5mins. Next, 500µl of freshly prepared 0.2M NaOH/1% SDS was added, mixed thoroughly by vortexing and left on ice for 5mins. This step is to lyse the bacterial cell wall and denature the DNA. To neutralise, 250µl of 3M KAc (pH 4.8) was then added, mixed thoroughly as before and left on ice for 10 mins.

The tube was then centrifuged twice for 10 mins at RT at 11,000rpm and in each case the supernatant was retained. This is to remove the cellular debris. RNase A (Sigma) was then added at a final concentration of 20µg/ml and left at 37°C for 10mins. The supernatant was then extracted twice with

200µl chloroform (Fisons), mixing the layers thoroughly by vortexing and then separating them by centrifugation for 1min.

The DNA was then precipitated by adding an equivalent volume of 100% isopropanol (Fisons) and immediately centrifuging for 10 mins at RT. The supernatant was discarded and the pellet then washed briefly in 500µl 70% ethanol and dried under vacuum.

2.1. 6 Precipitation with polyethylene glycol (PEG)

The DNA prepared for ABI sequencing was subjected to a further precipitation. The pellet was resuspended in 33.6µl dH_20, then 6.4µl NaCl and 40µl autoclaved 13% PEG_{8000} were added. After thorough mixing, the tube was incubated for 20 mins on ice and then the DNA was precipitated by centrifugation for 15 mins at 4°C. The supernatant was then removed and the pellet dried under vacuum. The pellet was then resuspended in 20µl of dH_20. To check the concentration of plasmid DNA recovered, 1µl of the 20µl was run on a 0.8% agarose gel, and the intensity of the bands compared with standards.

2.1.7 Preparation of plasmid DNA (large scale)

The selected bacterial colony was grown overnight at 37°C, with shaking at 250rpm, in 4ml Terrific broth plus ampicillin, then this 4ml was used to inoculate a 400ml for a further overnight culture. The culture was then centrifuged in 50ml volumes at 4000rpm for 20mins at 4°C. The pellet was then resuspended in 16ml GTE and left for 5mins on ice. Next, 32ml of freshly prepared 0.2M NaOH/1% SDS was added, mixed thoroughly by vortexing and left for 5mins on ice, before adding 16ml 3M KAc (pH 4.8), again mixing and leaving for 10 mins on ice.

The tubes were then centrifuged at 4000rpm for 20 mins at 4°C and the supernatant retained. This supernatant was then filtered gently through 4

layers of gauze into fresh centrifuge bottles. 36ml of isopropanol was then added and the bottles spun immediately at 4000rpm for 10 mins at 4°C. The supernatant was discarded and the pellet was washed with 20ml of 70% ethanol and centrifuged at 4000rpm for 10 mins at 4°C. The supernatant was once again discarded and the bottles drained. The pellet was resuspended in 2.5ml TE (pH 8) and transferred to a Universal tube.

Next, 3.55g CsCl and 200µl of EtBr (5mg/ml stock) (Boehringer Mannheim) were added and the solution transferred to a 'quickseal' tube (Beckmann). A balance tube was made up if necessary with TE, CsCl and EtBr. The tube was then spun for 20hrs at 80,000rpm at 20°C in a fixed angle rotor. The plasmid band was then carefully extracted through the side of the tube with a needle and syringe and placed in a 15ml tube. The plasmid DNA was then extracted three times with a 0.25x volume of CsCl-saturated isopropanol. Water was then added to increase the volume approximately twice, and then the DNA was ethanol precipitated. The resulting pellet was resuspended in 1ml water and the concentration of plasmid DNA checked by running 1µl on a 0.8% agarose gel and comparing the band intensities with standards.

2.2 Cloning of DNA Molecules into Plasmid Vectors

2.2.1 Strain of bacteria used

The competent cells were made from *E.coli*, strain XL1-Blue (Stratagene). The genotype of XL1-Blue cells is: recA1 endA1 gyrA96 thi-1 hsdR17 supE4 relA1 lac [F' proAB $lacI^q$ ZΔM15 Tn10 (Tet^r)]

2.2.2 Preparation of competent cells

Using a sterile loop, XL1-Blue bacterial cells from a frozen stock were streaked out onto a fresh L-agar plate and grown at 37°C overnight with continuous shaking at 250rpm. One of the resulting colonies was then used to inoculate 10mls of Terrific broth plus tetracycline, which was left at 37°C in

a shaking incubator overnight. This whole culture was then used to inoculate 1l of Terrific broth, and this was grown shaking at 37°C. The absorbance at 600nm of the culture was continually checked on a spectrophotometer.

The cells were harvested when the Abs_{600} was between 0.5 and 1.0, which is still in the log phase of growth. The flask containing the culture was then chilled on ice for 15-30 mins then the contents were centrifuged in ice cold 50ml Falcon tubes for 15 mins at 4000rpm at 4°C. The supernatant was discarded and the pellets resuspended in a total of 1l of sterile water before centrifugation as before. The pellets were resuspended in a total of 500ml sterile water and re-centrifuged. Pellets were then resuspended in 20ml of 10% glycerol and centrifuged again. The final resuspension was in 2-3ml of 10% glycerol. This cell suspension was then aliquoted into 40µl volumes in 1.5ml eppendorf tubes and snap frozen on dry ice plus methanol (Fisons). The aliquots were stored at -70°C.

2.2.3 Test transformation of competent cells

In order to quantify the competence of the cells prepared as above, they were subjected to a test transformation (see section 2.2.4) using 1ng of pBS vector alone as the transforming DNA. Cells transformed with pBS will contain an ampicillin resistance gene and can therefore be selected for on L-agar plus ampicillin plates. The competence of the cells is measured in cfu's per µg of transformed DNA. Values of 10^7 to $5x10^8$ cfu's/µg pBS were routinely achieved.

2.2.4 Plasmid vector

The plasmid vector used was pBluescribe (pBS; Stratagene) which is a 2746bp plasmid derived from pUC19 (Yanisch-Perron *et al*, 1985). It contains an ampicillin resistance gene and a multiple cloning site flanked by T3 and T7 promoters.

All the DNA molecules cloned possessed cohesive termini, which were compatible with a restriction enzyme site in the multiple cloning site of pBS. The digested insert molecules were therefore directly ligated into pBS linearised with the appropriate enzyme. However the 5'-phosphate groups from the vector were firstly removed by phosphatase treatment, so as to prevent recircularisation of the plasmid during ligation. The circular plasmid containing an insert of DNA was then used to transform competent cells.

2.2.5 Electro-transformation of competent cells

An aliquot of competent cells was thawed on ice, and then the transforming DNA (usually 1µl of the ligation reaction) was added, mixed with the cells and left on ice for 1min. This was then transferred to an ice-cold cuvette (Flowgen) and subjected to a pulse of 2.5kV in a BioRad Gene Pulser. The cells were swiftly mixed with 1ml of Terrific broth, transferred to an eppendorf tube and left at 37°C for 1hr to enable the cells to begin to express the ampicillin resistance gene conferred by the transformed plasmid. Aliquots of several different volumes were then spread onto L-agar plus ampicillin plates and incubated overnight at 37°C.

2.2.6 Selection for colonies that contain recombinant plasmids

Selection was by a simple process of blue/white selection. The polylinker of the vector pBS interrupts the β-galactosidase producing gene, lac Z. However the disruption is in-frame and results in a harmless insertion of a few amino acids into the β-galactosidase gene. Expression of this gene within bacterial cells can result in the production of a blue colour if the medium on which the cells are grown contains the chromogenic substance X-Gal and the derepressor of the Lac operon, IPTG. If however the polylinker site in the plasmid is interrupted by an insert of foreign DNA, a functional β-galactosidase enzyme cannot be transcribed and hence the resultant colonies are white.

2.3 Yeast Cell Culture and DNA Extraction

2.3.1 Media and solutions

All media were sterilised by autoclaving

AHC broth and agar

AHC is a rich selective medium which lacks uracil and tryptophan. It was used for selective growth of YAC recombinants prior to the isolation of YAC DNA and the production of plugs.

In 1litre water dissolve 1.7g yeast nitrogen base (without amino acids and NH_4SO_4) (Difco), 5g NH_4SO_4 and 10g casein hydrolysate (low salt). Adjust pH to 5.8. 17-20g of Bacto agar (Difco) were added to each litre of broth to make AHC agar. Autoclave. Add 50ml filter sterilised 40% glucose and 10ml of 2mg/ml adenine sulphate.

YPD broth and agar

YPD medium was used for growing yeast cultures.

In 1 litre water dissolve 20g bactopeptone (Difco), 5g yeast extract (Difco) and 5g NaCl. Add 1.5% agar for YPD agar. Autoclave. Add 40% sterile glucose.

SCE

In 500ml water dissolve 1M sorbitol, 0.1M Na_3Ci, 60mM EDTA, pH7.

Solution 1

6ml 50mM EDTA, 2ml SCE, 100µl β-mercaptoethanol (Sigma), 4mg zymolyase (ICN Biomedicals).

Solution 2

50ml 0.5M EDTA, 250µl 2M Tris.HCl, pH 7.5, 3.75ml β-mercaptoethanol.

Solution 3

200ml 0.5M EDTA, 1ml 2M Tris.HCl pH7.5, 2g N-laurylsarcosinate. Immediately before use add 1mg/ml proteinase K (Boehringer Mannheim).

2.3.2 Preparation of yeast DNA

A 5ml yeast cell culture was grown overnight in AHC and was centrifuged for 4 mins at 3000rpm (in a Sorvall T6000). The pellet was resuspended in 0.5ml 1M sorbitol, 0.1M EDTA solution, pH7.5 and transferred to a microfuge tube. 20µl of 2.5mg/ml zymolyase was added to break down the cell wall, and the tube incubated at 37°C for 2 hours. The tube was then centrifuged for 1 min at 11,000rpm, the supernatant discarded and the pellet resuspended in 0.5ml of 50mM Tris.HCl pH7.4, 20mM EDTA.

100µl of 10% SDS was added, and mixed in carefully but thoroughly, before incubation at 65°C for 30 mins. The proteins were then precipitated by adding 200µl 5M KAc, pH5.4. The tube was then left on ice for 1hr, then centrifuged at 100,000rpm for 5 mins. The supernatant was transferred to a fresh tube and to it was added 1 volume of isopropanol. This was mixed and left at RT for 5 mins, before centrifuging at 11,000 rpm for at least 10 secs. The supernatant was then removed with a pastette, and the pellet left to air dry. The pellet was then resuspended in 0.3ml TE, pH 7.4. RNA was removed by digesting with 1.5µl 10mg/ml pancreatic RNase (Sigma) at 37°C for 30 mins. The DNA was then precipitated by adding 15µl 5M NaCl and 900µl 100% ethanol, and then pelleted by centrifugation. The supernatant was then removed, the pellet was washed in 500µl of cold 70% ethanol and dried. The pellet was then resuspended in 200µl TE, pH7.4.

2.3.3 Preparation of yeast DNA plugs for PFGE

This method allows for the extraction of intact yeast chromosomes, including yeast artificial chromosomes (YACs), from cells. It was adapted from Maule (1994).

A 12ml culture was grown from a single colony. The culture was grown in selecting AHC medium at 30°C for 24 hours or until the cells reached approximately 10^8 cells/ml. The culture was then subjected to centrifugation at 3,000 rpm for 10 mins. The supernatant was discarded and the cell pellet washed twice in 15ml 50mM EDTA, pH7.5. After washing, the cell pellet was resuspended in 0.6ml Solution 1, and was mixed, but not vortexed. Immediately, 0.75ml 1% LMP agarose (Gibco, BRL) at 50°C in 0.125M EDTA, pH7.5 was added. After mixing thoroughly, 100µl aliquots of the solution were quickly dispensed into plug moulds and left to cool at 4°C. Once set, the plugs were ejected into a Falcon tube containing 10ml Solution 2 and were incubated at 37°C for 6 hours or overnight. The Solution 2 was then replaced with 10ml Solution 3 and the tube left at 50°C overnight. This step was repeated with fresh Solution 3. The Solution 3 was replaced with 0.5M EDTA and the plugs stored at 4°C.

2.4 Purification and Concentration of DNA

2.4.1 Phenol/chloroform/ether extraction

To remove proteins from DNA samples, the DNA can be extracted with organic solutions. Two methods were used:

a) DNA was extracted twice with a phenol/water/chloroform mixture (Applied Biosystems). This involved adding 1/10 volume of the organic mixture, vortexing vigorously and then centrifugation at 11,000rpm for 2 mins. The upper, aqueous layer was then carefully removed, avoiding precipitated protein at the interface between the two layers. The DNA was then recovered by ethanol precipitation.

b) DNA was extracted first with a 1/10 volume mixture of chloroform (Fisons) and isoamylalcohol (Sigma) in a 24:1 ratio. A second extraction was then performed with an equal volume of water-saturated ether (Fisons). After

separation of the aqueous layer from the ether, the tube was left open to the air for 10 mins to allow evaporation of any residual ether, and then DNA was ethanol precipitated.

2.4.2 Ethanol precipitation

To concentrate DNA and remove salts, a 1/10 volume of 2M NaAc, pH 5.5 was added to the DNA solution, followed by 2-2½ volumes of 100% ethanol at -20°C. The contents of the tube were mixed and then chilled at -70°C for 1hour to overnight. The tube was then centrifuged at 11,000rpm for 15 mins at 4°C. The supernatant was poured off, and the pellet dried under vacuum. The pellet was then resuspended in the desired volume of dH_2O.

2.4.3 Drop dialysis

DNA solutions were drop dialysed to remove salts. A 0.025µm drop dialysis filter (Millipore) was placed on the surface of a pool of sterile water in a Petri dish. After 5 mins, the DNA solution (maximum 20µl) was placed as a drop onto the surface of the filter and left for 30 mins, during which the concentration of salts in the sample equilibrates with the water. The drop was then 'sucked back up' into a clean tube.

2.5 DNA electrophoresis

2.5.1 Electrophoresis solutions

<u>20x TBE</u>

1M Tris.HCl, pH8.0; 20mM EDTA; 1M boric acid, pH8.3.

<u>20x TAE</u>

0.8M Tris.HCl, pH 8.0; 20mM EDTA; 0.4M acetic acid.

<u>10x DNA Loading Buffer</u>

20% Ficoll (Pharmacia), 100mM EDTA, orange G (Sigma).

6% Denaturing Polyacrylamide

Pre-prepared acrylamide/ bis acrylamide (19:1) sequencing solution (Severn Biotech Ltd.) was used, containing 6% w/v acrylamide, 7M urea and 1x TBE. 33mls of this solution was used to pour each gel. 33µl of TEMED (Sigma) and 0.2ml of freshly made 10% ammonium persulphate were added to the acrylamide before pouring the gel.

10x MOPS

0.2M MOPS, 0.05M NaAc, 0.01M EDTA, pH to 7.0, autoclave.

2.5.2 Size markers used in Gel Electrophoresis

250ng of the appropriate size marker was used per gel. Size markers used were:

a) λ DNA digested with *Hind III* (Boehringer Mannheim)

b) ϕX174 digested with *Hae III* (Boehringer Mannheim)

c) 1kb ladder (Gibco BRL)

Saccharomyces cerevisiae strains YP148 (Jones *et al*, 1989) and AB792 (Link and Olson, 1991) were used as size markers for PFGE.

Table 2.1 lists the sizes of the bands in λ*Hind III*, ϕX *Hae III* and the 1kb ladder, also the sizes of the chromosomes of YP148 and AB972.

Table 2.1

λ *Hind III* (kb)	ϕX *Hae III* (bp)	1kb ladder (bp)	YP148 (kb)	AB972 (kb)
23	1353	12216	2500	1640
9.5	1078	11198	1500	1105
6.5	872	10180	1125	955
4	603	9162	1035	930
2.3	310	8144	1000	830
2.0	281	7126	940	790
0.5	271	6108	830	750
	234	5090	790	690
	194	4072	750	585
	118	3054	680	445
	72	2036	600	350
		1636	550	285
		1018	440	240
		517	350	
		506	270	
		396	210	
		344	90	
		295		
		220		
		201		
		154		
		134		
		75		

2.5.3 Agarose gel electrophoresis

DNA molecules were separated according to their size on horizontal agarose medium EEO (Sigma) gels. The percentage of agarose used to make the gel depended on the size range of the DNA molecules to be resolved. Digested genomic DNA or plasmid DNA was commonly run on 0.8% agarose gels, whereas smaller fragments, such as most PCR products, were run on 1-2% agarose gels. All agarose gels were made with and run in 1x TAE. To stain the DNA, ethidium bromide was added to all agarose gels and the running buffer at a concentration of 250μg/ml of buffer. 1/10 of the sample volume of 10x loading buffer was added to the DNA prior to loading the sample on the gel. Either mini gels (30ml agarose) or midi gels (120ml agarose) were used depending on requirements; gels were run at 25-80V depending on resolution and run-time required.

DNA fragments were visualised on a mid range UV transilluminator and photographed using a video copy processor (Mitsubishi).

2.5.4 Purification of DNA from agarose

DNA fragments were run in low melting point agarose (Ultrapure LMP agarose, Gibco BRL) in 1x TAE. Gels were viewed on a mid range UV transilluminator and the required fragment was excised using a sterile scalpel blade as quickly as possible to minimise nicking of the DNA. Care was taken to ensure that a minimum of agarose was excised with the required DNA band.

Isolation of DNA from gel slices was by agarase treatment. The gel slice was diluted with water to decrease the percentage agarose to 0.8%, then 1/25 volume of 25x agarase buffer (Boehringer Mannheim) was added. Large gel slices were split into aliquots of ~100μl each. The gel slices were then melted at 70°C for 10 mins, during which they were mixed and spun down several times. The slices was left to cool to 37°C, then 1μl (0.5-1unit) agarase

(Boehringer Mannheim) per 100µl was added. The reaction was left to proceed for 2hrs at 37°C, then the DNA was ethanol precipitated at -70°C from 1hr to overnight.

2.5.5 Polyacrylamide gel electrophoresis

Denaturing polyacrylamide/urea gels were used to resolve DNA sequencing ladders. The following electrophoresis procedure was used for non-automated sequencing of PCR products.

Glass plates were prepared by washing thoroughly in detergent and distilled water, followed by 100% ethanol. The front plate was coated with dimethyldichlorosilane and left to dry. The back plate was treated with 30µl γ-methacryloxypropyl- dimethoxysilane (Sigma) and 30µl glacial acetic acid (Fisons) in 10ml 100% ethanol and left to dry. The plates were then rinsed briefly in 100% ethanol (back plate) or distilled water (front plate) and dried. Spacers were placed along the sides of the back plate, the front plate was placed on top, then the sides and bottom of the plates were sealed together with tape. Freshly prepared polyacrylamide was then poured between them. An inverted sharks-tooth comb was inserted to form a flat top to the gel, and clamps were applied to the plates' edges. The gel was allowed to set for at least an hour at room temperature.

Prior to running the gel, the tape was removed from the bottom of the plates. A metal plate was clamped to the front of the gel to ensure even heat distribution. The comb was removed and the surface of the gel flushed with 1x TBE to remove excess urea. The comb was then re-inserted and the gel loaded with up to 10µl sample. The gels were run in 1x TBE on a vertical slab gel apparatus (Scotlab) at 30W for 2 hours, or until the dye in the loading buffer had reached the bottom of the gel.

After electrophoresis, tape at the sides of the plates was removed and the front plate was lifted off, leaving the gel stuck to the back plate. The gel was fixed by washing in 15% methanol, 5% glacial acetic acid for 20 mins. It was then dried in an 80°C oven. The gel was then exposed to x-ray film.

2.5.6 Pulsed field gel electrophoresis

Agarose plugs were washed once in TE and equilibrated to the running buffer (0.5x TBE) over three hours. A 1% agarose (MP agarose, Boehringer Mannheim) gel was made up in 0.5x TBE and the plugs inserted into the wells. The wells were sealed with 1% LMP agarose and the gel run in pre-cooled sterile 0.5x TBE at 12°C. Gels were run using the Chef-DR II pulse field tank and control module (Biorad); run times and conditions were varied depending on the size resolution required.

After electrophoresis, DNA was stained by gently agitating the gel in running buffer containing 250μg/ml ethidium bromide for 20 minutes. The DNA was then visualised and photographed as in section 2.4.3.

2.5.7 Electrophoresis of RNA

RNA molecules were separated by electrophoresis on agarose gels containing 10% 10x MOPS and 17% formaldehyde. 1.2% medium EEO (Sigma) agarose was dissolved in 1x MOPS, left to cool to 50°C, then the formaldehyde was added and the gel poured immediately. The gels were left to set, and then loaded, in a fume hood.

The RNA sample to be loaded onto the gel (5μg) was made up to the required volume containing 50% formamide, 10% 10x MOPS and 16% formaldehyde. Prior to loading the sample was heated to 65°C for 20 mins, cooled on ice for 5 mins, and then 1x loading buffer was added. Gels were run at 30V overnight.

After electrophoresis, the RNA was stained by gently agitating the gel in dH_20 containing 250µg/ml ethidium bromide for 10 mins. The background staining was then removed by similarly agitating the gel in dH_20 for as long as necessary. The RNA was then visualised and photographed as for DNA.

2.6 Transfer of DNA and RNA to membranes

2.6.1 Southern transfer

DNA was transferred from gels to nylon membranes by capillary blotting. This method was adapted from Southern (1975).

Gels were photographed next to a ruler to allow for future sizing of DNA fragments.

To aid the transfer of large yeast chromosomes and YACs, pulsed field gels were additionally UV-irradiated on a mid range transilluminator for 3 mins, or gently agitated in 0.25M HCl at RT for 20 mins.

The DNA was denatured by gently shaking the gel in denaturing solution (0.5M NaOH, 1.5M NaCl) for 45 minutes. The gel was then neutralised by gently shaking in neutralising solution (1M Tris.HCl, 2M NaCl, pH 5.5) for 45 minutes and then rinsed. A large strip of 3MM filter paper (Whatman) was soaked in 20 X SSC (3M NaCl, 0.3M $Na_3C_6H_50_7.2H_2O$, pH7.0) and placed on a board. The ends of the paper were placed in a reservoir of 20x SSC, forming a wick. The gel was placed inverted on top of the wet filter paper, then a correctly sized piece of nylon membrane (Hybond-N, Amersham), was placed directly onto the gel. Two pieces of 3MM blotting paper, pre-soaked in 2x SSC, were placed on top of the membrane. Air bubbles were removed carefully. Any exposed wick was screened off with Saran wrap (Dow Chemical Company), then a weighted stack of paper towels was placed on top.

Pulsed field gels were blotted for two days, with a change of paper towels midway. Other gels were blotted for either 5 hours or overnight. After blotting, the membranes were left to air dry, and then DNA was bound to the filter by exposing it to 1200μJoules of UV irradiation in a Stratalinker (Stratagene). Alternatively the membranes were baked at 80°C for 1hour. Nylon membranes were stored in Saran wrap or between two sheets of 3MM paper at RT.

2.6.2 Transfer of RNA to nylon filters (Northern blotting)

RNA was transferred from agarose gels to Hybond-N nylon membrane in the same way as for DNA except that the denaturation and neutralising steps were omitted, and 10x SSC replaced 20x SSC as the solution beneath the wick.

2.6.3 Transfer of bacterial colonies to filters and replication of filters

Nitrocellulose (Schleicher and Schuell) or nylon (Hybond-N, Amersham) circular filters were labelled with an asymmetric pattern of dots and laid carefully on the surface of the agar plate, so as to avoid air bubbles. The filter was left on the surface for approximately 10 secs while the pattern of orientation dots from the filter was copied onto the bottom of the Petri dish. The filter was then peeled off, being careful not to smear the colonies, and left face up. To make a replica, another, unmarked, filter was then gently laid on top of the first filter and pressure was applied evenly across the surface to transfer the colonies. The pattern of dots was copied onto the back of the replica filter before separating the two. The colonies on the agar plates were left at 37°C for several hours to 'recover'.

2.6.4 Fixation of the bacterial colonies to filters by lysis

One sheet of 3MM paper was soaked in each of the following solutions:

a) 10% SDS

b) denature - 1.5M NaCl, 0.5M NaOH

c) neutraliser - 1.5M NaCl, 0.5M Tris.HCl (pH 7.5)

d) 2x SSC

The excess was then poured off and the sheets of 3MM placed in trays. The filters were placed face up on the sheets of 3MM and left for 1 min in SDS, then 3 mins in denature, 3 mins in neutraliser and 30 secs in SSC. The filters were left to air dry and were then baked at 80°C in a vacuum for 20 mins (nitrocellulose) or UV irradiated as in section 2.6.1 (nylon).

2.7 Radiolabelling of DNA

2.7.1 Preparation of DNA for probes

DNA to be used as a probe was either in solution or contained within a gel slice. 25-50ng was labelled for a single hybridisation. Gel slices to be used as hybridisation probes were treated as follows. After electrophoresis of DNA in an LMP gel the required fragment was excised and diluted with 1-3 times (depending on concentration) its volume of water. The slice was melted at 70°C and vortexed, before storage at -20°C. Before use in a labelling reaction, the slice was remelted at 70°C.

For oligonucleotide probes, 25-50ng of the precipitated oligonucleotide was used.

2.7.2 Random Priming of DNA probes

This method is adapted from Feinberg and Vogelstein (1983 and 1984).

A labelling reaction with [α-^{32}P]-dCTP involves random priming from hexanucleotides and then polymerisation along the DNA strand catalysed by the Klenow fragment of *E.Coli* polymerase 1. A radiolabelled base is incorporated at every C nucleotide.

The DNA strands were firstly denatured by heating to 100°C for 10 mins. DNA in solution was then kept on ice to prevent reannealing of the strands, DNA in a gel slice was cooled to 37°C. The labelling reaction was carried out using either a Random Prime kit (Boehringer Mannheim) or a Hi-Prime kit (Boehringer Mannheim). For the random prime kit, 11µl total of DNA plus water was mixed with 3µl 10µCi/µl [α-^{32}P]-dCTP (Amersham), 1µl (2 units) Klenow enzyme, 1µl each of dATP, dTTP and dGTP and 2µl reaction buffer. For the Hi-Prime kit, 13µl of DNA plus water was mixed with 3µl 10µCi/µl [α-^{32}P]-dCTP and 4µl Hi-Prime (which contains the enzyme, nucleotides and buffer). The reactions were then incubated at 37°C for 1 hour (Random Prime or Hi-prime with a gel slice) or 10 mins (Hi-Prime with DNA in solution).

The percentage incorporation of the radiolabelled nucleotide was checked by TCA precipitation of ~0.5µl of the reaction mix on a GF/B circular filter (Whatman). If the incorporation was 50% or above, the unincorporated nucleotides were removed by running the probe through a Sephadex G-50 Nick column (Pharmacia Biotech). The storage buffer was removed and then the column was washed by running through ~600µl TE. The probe was then added to the top of the column, followed by 380µl TE. The probe was then eluted with a further 400µl TE, and collected in a tube containing 500µg sonicated salmon sperm DNA (Sigma). The probe was denatured by heating to 100°C for 10 mins, and was kept on ice until adding it to the prehybridisation solution in the bottle.

Hybridisation due to high copy number repetitive elements in the probe could be minimised by 'stripping' the probe. The probe was denatured at 100°C for 10 mins in the presence of 500µg sonicated salmon sperm DNA and 250µg sonicated human DNA. It was then left to reanneal for ½ hour at 68°C before addition to the hybridisation bottle.

2.7.3 End-labelling of DNA oligonucleotides

This labelling reaction involves the transfer of the radiolabelled terminal phosphate group of [γ-^{32}P]ATP to the terminal 5'-OH group of the oligonucleotide.

25-50ng oligonucleotide DNA was mixed with 3µl (30µCi) [γ-^{32}P]ATP (Amersham) and 1µl (10 units) PNK (Boehringer Mannheim), in a total of 20µl 1x PNK buffer (5mm Tris.HCl, pH 8, 1mM $MgCl_2$, 0.5mM DTT). The reaction was incubated at 37°C for 40 mins, then the probe was added to the prehybridisation solution in the bottle.

2.8 Hybridisation Protocols

2.8.1 Solutions

<u>(Pre)hybridisation solution</u>

5x SSC, 10% dextran sulphate, 0.1% sodium pyrophosphate, 0.1% SDS and 5x Denhardt's solution. 5x Denhardt's solution contains 0.1% ficoll, 0.1% PVP and 0.1% BSA (Sigma). Filter before use.

<u>Oligonucleotide hybridisation solution</u>

5x SSC, 0.05% BSA, 0.05% ficoll, 0.1% SDS, 0.05% PVP, 0.1% sodium pyrophosphate. Filter before use.

2.8.2 Prehybridisation of filters

Under a solution of 2x SSC, the filter was placed between two slightly larger sheets of gauze. For hybridisation to multiple filters with a single probe, up to 3 layers of filters could be placed in the same bottle, with layers of gauze separating them. Air bubbles trapped between the filter and the gauze were removed. The filter and gauzes were then rolled up together, transferred to a glass hybridisation bottle (Hybaid) and unrolled onto the surface of the bottle.

Denatured sonicated salmon sperm DNA was added to the prehybridisation solution at a concentration of 100μg/ml. 12ml of prehybridisation solution was added to a small bottle and 18ml to a large bottle, and the bottle rotated in an hybridisation oven (Hybaid) at 68°C for a minimum of 1 hour before adding the probe.

For oligonucleotide probes, prehybridisation (and hybridisation) was performed at a temperature ~10°C lower than the T_m of the oligonucleotide, commonly 48°C.

2.8.3 Hybridisation conditions

For random-primed probes, hybridisation was carried out at 68°C overnight. For oligonucleotide probes, the hybridisation was overnight at a temperature 10°C lower than the T_m of the oligonucleotide (commonly 48°C).

2.8.4 Washing conditions

Following hybridisation, the filters were removed from the bottle and separated from the layers of gauze. In a plastic tray, the filters were washed with 3 changes of ~500ml washing solution (0.1-4x SSC, 0.1% SDS) using gentle agitation. The temperature of the wash and the concentration of SSC depended on the washing stringency required. For random-primed probes, a high stringency wash (3x 20 mins in 0.1x SSC, 0.1% SDS at 68°C) was commonly used. For oligonucleotide probes, the washing steps were far less stringent (3x 5 mins in 4x SSC, 0.1% SDS at RT).

The filters were then wrapped in Saran wrap (Dow Chemical Company), avoiding creases.

2.8.5 Detection of hybridisation signal

Autoradiography The filters were placed in a light-tight cassette with a signal enhancing screen. They were then exposed to X-OMAT x-ray film (Kodak) for a length of time dependent on the amount of radiolabelled probe left bound to the filter (several minutes to several days). Filters hybridised to ^{32}P-labelled probes were exposed at -70°C, those labelled with ^{35}S were exposed at RT. Stratagene Glogos II luminescent markers were used for alignment. The film was developed on an automatic x-ray film processor RGII (Fuji).

Phosphorimaging Alternatively, the filters were exposed to a phosphor screen (Molecular Dynamics) for hours to several days. The screen was then scanned on a PhosphorImager (Molecular Dynamics), where a laser beam converts the radioactive signal into a digital image, with variations in the pixel value proportional to the amount of radioactive signal present. The grey-scale image was adjusted as desired and was then printed on a laser printer.

2.8.6 Removal of radioactive probe from filters

Filters may be used several times with different hybridisation probes, so it was sometimes necessary to remove radioactive probes from filters. This was done in one of two ways. Firstly, the filters could be washed in 0.4M NaOH at 50°C for 30 mins, to denature the probe DNA strand from the surface of the filter. The neutral pH was then restored by a 30 min wash in 0.1x SSC, 0.1% SDS, 0.2M Tris.HCl, pH 7.5, also at 50°C. Alternatively the filters could be subjected to one very stringent wash, in 0.1% SDS (i.e. 0x SSC) at 100°C for 5 mins.

The filters were then exposed to x-ray film overnight to check that all the probe had been removed.

2.9 Enzymatic Manipulation of DNA

2.9.1 Restriction enzyme digestion of genomic DNA

Digestions of DNA with restriction endonucleases were carried out in the appropriate buffer at the recommended temperature. Restriction enzymes were supplied by Boehringer Mannheim, NEB and Gibco BRL. 1μg of DNA was digested in 5-10μl containing 1/20 volume 0.1M spermidine (Sigma), using 1-2 units of enzyme for 1½ hours. Human genomic DNA was digested overnight. If two different enzymes were used, both of which required the same buffer, the digests were carried out simultaneously. Otherwise, after digestion with one enzyme, the sample was drop dialysed for 30 mins. The appropriate buffer was then added and the second digestion carried out. When necessary, reactions were terminated by heating to 68°C or 80°C for 15 minutes, according to the heat sensitivity of the enzyme (NEB Catalogue, 1992). Restriction digests which were to be run on gels were terminated by the addition of 1/10th of their volume of "stop mix" (100mM EDTA, pH8.0; 20% Ficoll and Orange G).

2.9.2 Restriction digestion of agarose plugs

This protocol was adapted from Maule, 1994.

Plugs were soaked in a large excess of sterile TE at room temperature for 10 mins. A 20mg/ml solution of PMSF was prepared, at 55°C, in propan-2-ol. Plugs were then transferred to a solution of TE containing 40μg/ml PMSF (Sigma). PMSF is a potent inhibitor of Proteinase K and was used to inactivate residual Proteinase K that may be present in the plugs as a result of their preparation. The plugs were incubated in this solution for 30mins at 50°C with one change of solution after 30 minutes. The plugs were then soaked in 30 plug volumes of 1x restriction enzyme buffer for 2 hours at room temperature, with frequent inverting.

For restriction enzyme digestion, each plug was placed in an eppendorf tube and put on ice. Plugs were digested in 100µl volume which contained 1x restriction buffer, 0.1% Triton X100 (Sigma), 200µg/ml BSA and 20 units of restriction enzyme. Any air bubbles were removed to ensure good contact of the plug with the enzyme solution. Digestion was carried out overnight at the appropriate temperature. Following digestion, the tube was cooled on ice for 15 minutes, and the liquid removed with a fine tipped pastette. The plugs were then rinsed in 1ml ice-cold TE. The TE was removed, then 200µl 'plug stop mix' (9.4ml dH_20, 0.4ml EDTA, pH 8, 0.25ml TAE and Orange G) was added and the tubes were left on ice for 20 mins.

2.9.3 Dephosphorylation of 5' termini

Calf intestinal phosphatase (CIP) was used to dephosphorylate the 5' ends of the vector molecules before cloning. This prevents recircularisation of vector molecules during the ligation step.

7µg DNA was dephosphorylated with 0.1 unit CIP (Boehringer Mannheim) in 50µl 1x CIP buffer (10mM Tris.HCl, pH 8.3, 1mM $ZnCl_2$, 1mM $MgCl_2$) at 37°C for 30 mins. The reaction was stopped by the addition of 1µl of 0.5M EDTA. The dephoshorylated DNA was then chloroform/isoamyl alcohol then ether extracted and ethanol precipitated. The pellet was resuspended in 10µl TE.

2.9.4 Phosphorylation of 5' termini

Addition of a phosphate group to a DNA strand ('kinasing') is by transfer of the terminal phosphate of γATP. It is necessary to kinase oligonucleotides before they can be ligated to restriction enzyme digested DNA.

4µg oligonucleotide DNA was mixed with 1µl (10 units) polynucleotide kinase (PNK; Boehringer Mannheim) and 0.5-1mM ATP (Boehringer Mannheim) in 1/10 volume 10x PNK buffer (50mM Tris.HCl, pH 7.5, 10mM $MgCl_2$, 5mM DTT) in a total volume of 50µl. The reaction was incubated at 37°C for 30

mins. The sample was then chloroform/isoamylalcohol then ether extracted and ethanol precipitated. A recovery of 25% was assumed.

2.9.5 Ligation of cohesive termini

On ice, the insert DNA and the vector DNA were mixed, typically in a ratio of 5:1 vector to insert. 10-100ng of vector DNA was normally used. The reaction was carried out in 10-50µl using 0.1 unit of bacteriophage T4 DNA ligase (Boehringer Mannheim). in 1x ligase buffer (66mM Tris.HCl, pH 7.5, 5mM $MgCl_2$, 1mM DTT, 1mM ATP). The ligation mixture was then incubated at 16°C overnight, or at RT for 4½ hours (the latter for vectorette and splinkerette ligations). The enzyme was then inactivated by heating to 65°C for 15 mins.

For a trial ligation, 100ng of bacteriophage λ digested with *Hind III* was religated under the same conditions. The extent of ligation could be visualised by running the religated product on a 0.8% agarose gel.

2.10 Oligonucleotides

2.10.1 Oligonucleotide Synthesis

Oligonucleotides were synthesised (by Agnes Gallagher) as ammonium stocks on an Applied Biosystems 381A oligonucleotide synthesiser. Oligonucleotides for use as FISH probes or for magnetic bead capture were synthesised incorporating a 5' biotin moiety.

Oligonucleotides were precipitated from ammonium stocks by ethanol precipitation of 350µl of the stock. The precipitated DNA was resuspended in 500µl water and the concentration assessed by measuring the Abs_{260}. An absorbance of 1 corresponds to a concentration of 25µg/ml for single stranded DNA.

2.10.2 Oligonucleotide primer Design

The vectorette primers were obtained from ICI. Splinkerette sequences and primers were designed to be comparable with those of the vectorette (personal communication). The sequences of two published primers were used, that is the primers '291' and '292', for the amplification of recombinant DNA inserts in the plasmid pBS ('forward' and 'reverse' primers from Stratagene catalogue, see table 2.2 for sequences).

Other primers were designed using the programme Oligo4 (Hybaid). Primers were between 19 and 46 nucleotides in length depending on the DNA source on which they would be used. They were chosen to be stable oligonucleotides with a T_m between 55°C and 70°C. A maximum hairpin structure or potential dimerism of 3 nucleotides (preferably less) was permissible, but only a 2 nucleotide internal match if it involved the 3' end. If possible the A+T : C+G ratio was near 50%.

In two instances, it was necessary to synthesise slightly degenerate primers due to unresolved sequence ambiguities.

2.10.3 Duplexing oligonucleotides

In certain cases it was necessary to form a duplex between two complementary (or partially complementary) oligonucleotides. The two oligonucleotides (150µg/ml each) were mixed together in 10µl 1x TM (10mM Tris.HCl, pH 7.5, 5mM $MgCl_2$) and the tubes were placed in a small volume of water at 90°C. The water was then left to cool to below 30°C over a period of 15-30 mins.

2.11 Amplification of DNA by the Polymerase Chain Reaction (PCR)

PCR is a technique used for the amplification of a specific DNA sequence (Saiki *et al*, 1988; Mullis and Faloona, 1987). The specificity is provided by oligonucleotide primers complementary to the 5' ends of the two strands of the sequence to be amplified. These primers anneal to the template DNA strand and direct amplification by a thermostable polymerase. The PCR reaction is a series of cycles, each consisting of three steps. The first is a short high temperature step (92-94°C) to denature the template. This is followed by a reduction in temperature to allow the primers to anneal to the template (~50-65°C depending on the primers used), then finally the temperature is increased to 72°C (or 68°C for long PCR) to permit extension from the primers by the polymerase. Multiple cycles of these three steps result in exponential amplification of the desired sequence.

2.11.1 PCR conditions

Commonly 100ng of template genomic DNA was amplified in 50µl in 0.5ml microcentrifuge tubes (Robbins Scientific), with 0.2µl (1 unit) Amplitaq (Perkin Elmer Cetus) in 1x reaction buffer (10mM Tris.HCl, pH 8.3, 50mM KCl, 1.5mM $MgCl_2$, 0.01% (w/v) gelatin) and 0.1% Triton X-100 (Sigma). The reaction employed 300ng of each primer and 200µM each dNTP (Advanced Biotechnologies). The reaction was overlaid with 40µl mineral oil (Sigma).

PCR programmes were run on an Hybaid Omnigene machine or Perkin Elmer Cetus DNA thermal cycler. The basic programme was 30 cycles of 3 different steps:

a) Denaturation: A short step at 93-94°C. Usually 30 secs in the first cycle and 15 secs thereafter.

b) Annealing: The annealing temperature is critical in determining the success of the PCR reaction, and is dependent on the structure of the

primer. Primer T_m's were determined using the 'Melt2' programme (J.G. Wetmur, personal communication); an annealing temperature of ~5°C lower than the T_m was used. Annealing steps were 30 secs or 1 min long.

c) Extension: Carried out at 72°C. The extension time is also critical in determining the success of the PCR, since longer products need longer extension times (approximately 1 min per kb of sequence). Ten cycles of three different extension times was generally used, the length progressively increasing until the time estimated for effective elongation was reached.

After the reaction, 5-10µl of the PCR reaction was visualised by electrophoresis on 0.8-2% agarose gels.

Steps to increase specificity

'Hot Start' PCR (Chou *et al*, 1992) was commonly used. The PCR reaction was set up as 2 separate mixes, one containing the DNA and primers, the other containing the enzyme, buffer and nucleotides. The mixes were heated to 90°C for several minutes, before combining them. Addition of the enzyme to the DNA at a temperature at which the DNA is denatured helps to prevent mispriming before the initial denaturing temperature is reached.

'Touch Down' PCR (Don *et al*, 1991) implies a decreasing annealing temperature per cycle until the final annealing temperature is reached. This means that during the first few critical cycles, only highly specific primer-template interactions can occur, which selects for amplification of the correct product during the rest of the reaction. An annealing temperature 10°C above the calculated value was used in the first cycle, which decreased by 2°C each cycle until the final annealing temperature was reached.

Long PCR

For amplification between the seven novel gene products resulting from the CSC experiment, an 'Expand' (Boehringer Mannheim) or 'Takara' (Takara

Shuzo Co.) long PCR kit was used. These kits incorporate two thermostable DNA polymerases, *Taq* polymerase for high processivity, plus a second polymerase which possesses a 3'-5' exonuclease (proofreading) activity. The kits were used according to the manufacturer's instructions, apart from slight changes to the suggested programme and the procedure for setting up the reactions. Details are given in section 6.5.3.

2.11.2 PCR amplification of plasmid inserts

This method was adapted from Taylor (1991).

A toothpick stab of a bacterial colony was transferred to a tube containing 50µl of water. This tube was then heated to 96°C for 4 mins. 1µl of this was then used as the template for a PCR reaction. Primers used were Universal primers designed to the polylinker sequence of pBS, 291 and 292 (see table 2.2).

The PCR reaction mix was exactly as in section 2.11.1.
The PCR programme used was:
Denaturing: 93°C for 30 secs in first cycle and 15 secs thereafter.
Annealing: 52°C for 1 min.
Extension: 72°C, 10 cycles of 1 min, 10 cycles of 2 mins, 10 cycles of 3 mins.

2.12 Sequencing of DNA

2.12.1 Cycle sequencing of plasmid inserts

Sequencing of plasmid clones was done on an ABI automated sequencing machine (Applied Biosystems). Thermal cycling of sequencing reactions increases signal intensity and decreases sensitivity to reaction conditions. Dye terminator chemistry was used, that is dideoxy-nucleotides terminated the sequencing reaction at each base to generate the ladder of fragments.

Each of the four dideoxy-nucleotides is labelled with a different dye, such that the attached computer can recognise the peak of dye corresponding to a particular base at each position.

Plasmid DNA for sequencing was prepared by alkaline lysis followed by PEG precipitation according to the protocols in sections 2.1.5 and 2.1.6. The sequencing reactions were performed using a Taq DyeDeoxy Terminator Cycle Sequencing kit (Applied Biosystems) according to the manufacturer's instructions. Ideally, approximately 200ng plasmid DNA was used in the reaction, although results were produced with as little as ~20ng. 18ng of either primer 291 or 292 was used per reaction, depending on the orientation of the sequence required.

The cycle sequencing reaction was performed on a Perkin Elmer Cetus DNA thermal cycler. The programme was as follows: Place tubes in rack preheated to 96°C. Perform 25 cycles of 96°C for 30 secs, 50°C for 15 secs and 60°C for 4 mins. Rapid thermal ramp to 4°C and hold.

As much as possible of the mineral oil was removed with a pipette, then 90µl water was added. The DNA was then extracted twice with an equal volume of a phenol/water/chloroform mixture (Applied Biosystems) and then ethanol precipitated, using Applied Biosytems NaAc and 3 volumes of cold 100% ethanol. The tubes were then placed on ice for 15 mins, before centrifugation for 15-20 mins at 11,000rpm. The supernatant was then removed, and the pellet dried for 1 min at 85°C in a thermal cycler. Immediately before loading onto a sequencing gel, each pellet was resuspended in 4µl stop mix (1µl of 50mM EDTA, pH 8, 30mg/ml Blue Dextran to 5µl deionised formamide).

The samples were loaded onto a 6% acrylamide gel (prepared by Agnes Gallagher), which was run in 1x TBE at 2500V, 45mA for 12 hours. The sequence data was gathered by the computer and displayed as a visual

output of differently coloured peaks, or DNA sequence which was directly amenable to computer analysis.

2.12.2 Sequencing of PCR products

This method has been modified from Winship, 1989.

In order to directly sequence PCR products, the required product must be purified from a gel in order to separate it from primer dimers and any spurious products. A substantial amount of DNA is required in the reaction itself, so in some cases the product of several PCR reactions was pooled and concentrated into a smaller volume by ethanol precipitation. The resulting product was then gel purified by running it on a 1% (typically) LMP agarose gel and excising the required band. The DNA was recovered from the gel slice by agarase treatment and ethanol precipitation. The precipitate was resuspended in 10-40µl and the DNA concentration checked by running 1µl on a gel next to a known amount of size marker. Approximately 200ng DNA was used per sequencing reaction.

The sequencing reactions were carried out using a Sequenase Version 2.0 kit (Amersham). Firstly, the DNA was mixed with 300ng sequencing primer, 1.25µl DMSO and 2.5µl 5x Sequenase reaction buffer and made up to a total volume of 12.5µl with water. The mix was then denatured by heating it to 96°C for 3 mins on a thermal cycler. The tube was then immediately plunged into liquid nitrogen or dry ice to snap freeze DNA mix and prevent the strands from reannealing.

A labelling mix was made up from the kit containing 1.75µl enzyme dilution buffer, 1µl DTT, 0.4µl dITP labelling mix and 0.25µl Sequenase enzyme per reaction. To this was added 1.6µl dH_2O and 0.5µl of 10µCi/µl [α-^{35}S]-dATP (Amersham) per reaction, and the tube kept on ice. One tube of DNA mix was thawed quickly by rubbing between the fingertips, then 5.5µl of the

labelling mix was immediately added and mixed by pipetting. 3.5µl of this was then immediately applied to the side of 4 eppendorf tubes, each one containing 2.5µl of one of the 4 termination mixes containing 10% DMSO (G,A,T and C). The 4 tubes were then briefly spun to combine the labelling and termination mixtures simultaneously. The time from thawing the DNA to mixing with the termination mix was approximately 45 secs.

The termination tubes were transferred to a 37°C water bath and incubated for 5 mins. The reaction was then stopped by the addition of 4µl kit stop mix (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF). Reactions were stored at -20°C until loading on a gel. Prior to loading they were heated to 80°C for 2 mins to denature the DNA.

2.12.3 Analysis of sequence data

The sequences obtained were edited and manipulated using GCG (Genetics Computer Group) programmes available through the HGMP Computing service (Devereux *et al*, 1984). Database searches were performed using the GCG Blast (Altschul *et al*, 1990) and Fasta (Pearson and Lipman, 1988) programmes.

2.13 'Catch Linkering' of YAC and cDNA

Catch linkering is a method for amplifying DNA, for instance a whole YAC, into a simplified resource. It is an alternative to Alu-PCR (Nelson *et al*, 1989), yet is not dependent on the presence of any specific sequence elements other than frequent restriction enzyme sites. DNA is digested with *Sau3AI*, and then a short double stranded 'catch linker' is ligated onto each cohesive end. The intervening DNA can then be amplified by PCR using primers designed to the catch linkers. Catch linkered material can be labelled and used as a probe for FISH (Shibasaki *et al*, 1995), or (as in this case) as a substrate for Coincident Sequence Cloning.

2.13.1 Preparation and digestion of DNA

YAC DNA was isolated from a pulsed field gel as a 1% LMP agarose slice, using minimal ethidium bromide staining and UV irradiation. The DNA was then purified from the gel slice by agarase treatment, and ethanol precipitated for 1 hr.

The resulting pellet was resuspended in 5µl 1x restriction enzyme buffer A (Boehringer Mannheim) and digested with *Sau3AI* at 37°C for 1 hr. The enzyme was heat killed by heating at 70°C for 15 mins. The sample was then placed on ice.

2.13.2 Preparation of catch linkers

Catch linker oligonucleotides were 477 / 479 for YAC DNA and 727 / 731 for cDNA (table 2.2).

Oligonucleotide 477 was synthesised with and without biotin at the 5' end. Oligonucleotides 479 and 731 were phosphorylated at the 5' end to enable them to be ligated to the restriction enzyme digested YAC DNA. 1µg of each oligonucleotide was duplexed in 10µl TM to form the double stranded linker.

2.13.3 Ligation of catch linkers

100ng linker DNA was then ligated onto ~20ng digested YAC (quantity judged from the initial pulsed field gel) or cDNA, in 20µl volume at 16°C overnight. The sample was then briefly spun and placed on ice.

2.13.4 PCR amplification

1µl of the ligated material was then amplified by PCR, using 600ng oligonucleotide 477 or 727 as the primer. Biotinylated 479 was used if labelled product was desired (see table 2.2).

The PCR programme used was:

Denaturation: 93°C for 15 secs

Annealing: 54°C for 30 secs

Extension: 72°C for 1 min (cycles 1-10), 2 mins (11-20) and 4 mins (21-35). 10μl of the PCR product was examined on a 1.5% agarose gel. The remainder was ethanol precipitated and resuspended as appropriate to future use.

2.13.5 Detection of biotin labelling efficiency

The extent to which the catch-linker PCR product was labelled with biotin could be detected by binding the biotin residues to streptavidin conjugated with the enzyme alkaline phosphatase. Action of the enzyme on a substrate (nitroblue tetrazolium, NBT), together with another compound (5-bromo-4-chloro-3-indolylphosphate, BCIP) causes formation of a coloured product.

1μl of a series of dilutions of the PCR product (1/10 to 1/1000) was spotted onto a nitrocellulose filter, and the filter baked at 80°C for 20 mins. The filter was then re-hydrated in Buffer 1 (0.1M Tris.HCl, pH 7.5, 0.15M NaCl) for 1 min, then incubated in Buffer 2 (3% BSA in Buffer 1) for 1 hour at 65°C in a sealed hybridisation bag. Immediately prior to use, a dilution of alkaline phosphatase-streptavidin (AP-SA, Vector Laboratories) conjugate was made by adding 20μl of the 1mg/ml AP-SA solution to 20ml Buffer 1 in a polypropylene tube. The filter was then incubated in the AP-SA solution for 10 mins at RT with gentle agitation. The solution was removed and the filter washed twice with 40-80ml Buffer 1 for 15 mins at RT with gentle agitation. An alkaline phosphatase substrate kit (Vector Laboratories), containing the NBT and BCIP, was then used according to the manufacturer's instructions to cause the formation of the coloured product. The filter was incubated with the dye solution in a sealed hybridisation bag for 30-270 mins in the dark. When suitable staining had developed, the reaction was stopped by washing the filter in 20mM Tris. HCl, pH 7.5, 0.5mM EDTA .

2.14 Coincident Sequence Cloning

Coincident Sequence Cloning (CSC) is a method for isolating the sequences shared between two DNA resources (see section 1.8.5). This is achieved by denaturing the two DNA sources, mixing them ('integration') and then recovering the duplexes which are formed between the two resources (inter-resource duplexes, IRDs). In this context, the two resources used were the YAC D0485 and foetal brain cDNA, so that the product should be transcribed sequences from the YAC.

At all times during the preparation of the Source DNAs, their integration and cloning, great care was taken so that the samples would not become contaminated. All materials used were dedicated solely to CSC, fresh solutions were made up and autoclaved, and all labware was washed with water and then ethanol.

2.14.1 Preparation of YAC DNA

The YAC D4085 was catch linkered (section 2.13) using biotinylated 477 as the PCR primer, by Tony Brookes.

Before use in the CSC experiments it was necessary to purify the catch linkered product from a gel, to remove primer dimers. It is preferable to do this without staining the DNA with ethidium bromide or exposing it to UV irradiation.

The catch linker PCR product was ethanol precipitated and resuspended in 10µl dH_20. 9µl of the sample was then run on a 1.5% LMP agarose gel in a minigel apparatus, without ethidium bromide staining. The remaining 1µl was run in a parallel, test track next to a ϕX *Hae III* marker track. When the loading dye had reached the bottom of the gel, this test and marker tracks was sliced away from the gel and stained with ethidium bromide. The DNA in these tracks was then viewed on a UV transilluminator. The top and bottom

positions of the catch linker product with respect to the marker were noted, and the gel was nicked where it should be cut to remove the primer dimers. This test track was then aligned alongside the rest of the gel, and the gel sliced across at the position of the nick.

In order to concentrate the product DNA into a smaller volume of gel, the orientation of the gel was reversed and the product run back up towards the wells. Loading dye placed in a hole at the bottom of the gel before turning it round was used to judge when the run was complete. A second ϕX *Hae III* marker track was sliced off and stained, and nicks made at the positions corresponding to the top and bottom of the catch linker product smear. This track was then aligned with the track containing the product, and the product was excised from the gel, guided by the nicks in the marker track.

The DNA was purified from the gel slice by agarase treatment and then was ethanol precipitated overnight. The resulting pellet was resuspended in 5µl dH_20. The concentration was estimated by running 0.5µl on a 1% agarose gel. The remaining 4.5µl (~800ng) was used in the HF-CSC experiment.

2.14.2 Preparation of cDNA

10µg cDNA was catch linkered (section 2.13) using 727 as the PCR primer, by Tony Brookes. For the EL-CSC experiment, the linkers were then removed by complete digestion with *Sau3AI*.

2.14.3 Preparation of 'blocking' DNA

Human Cot1 DNA and ribosomal DNA were used to preanneal to the denatured Source I and II DNAs, to prevent inter-resource hybridisation due to high copy repeat elements and ribosomal sequences.

Ribosomal Blocking DNA 50µg yeast DNA, plus 20µg each of the two ribosomal DNA plasmid clones pA and pB (obtained from Dr I. Gonzalis,

Hahnemann University, Philadelphia, USA) were mixed in 2ml dH_20 and then sonicated twice for 30 secs to reduce the fragment length to approximately 500bp to 1kb. The 2ml was aliquoted into eppendorf tubes and ethanol precipitated. The precipitate was resuspended in 45µl dH_20, and the concentration checked by running 1µl of a 1/10 dilution on a 1.2% agarose gel.

Cot1 DNA Two 10µg aliquots of Cot1 human genomic DNA were ethanol precipitated and vacuum dried.

4µg of the ribosomal blocking DNA in 3µl was used to resuspend each pellet of Cot 1 DNA.

2.14.4 Integration of the YAC and cDNA

800ng YAC DNA and 10µg cDNA were made up separately to 5.4µl in water. To each was added 10µg Cot 1 DNA and 4µg ribosomal blocking DNA in 3µl volume and 1.8µl 1M NaOH and the tubes placed at 37°C for 5 mins to ensure the denaturation is complete. The tubes were then warmed to 50°C and 9.8µl FNET/HCl (6µl formamide, 2µl NET (400mM Tris. HCl, pH7.8, 2.5M NaCl, 50mmM EDTA), 1.8µl 1M HCl) at 50°C was added to each to neutralise the alkali. The neutral pH of each was confirmed by spotting 0.5µl of each onto Indicator paper. The tubes were then incubated at 45°C for pre-annealing with the blocking DNA. The cDNA was incubated for 4hours, and the YAC DNA for 30 mins. After this, the two DNA mixes were combined and left submerged at 45°C overnight to permit the formation of the IRDs.

The integration and subsequent steps were also performed using controls of 'cDNA only' and 'no DNA'.

2.14.5 Selection of the Inter Resource Duplexes (IRDs) (HF-CSC)

IRD selection was by means of attachment of the biotin residues on the YAC DNA linkers to streptavidin coated magnetic beads. cDNA:cDNA duplexes and non-specific IRDs were removed by washing steps, whilst retaining the DNA attached to the beads. The strands present in a duplex with the biotinylated YAC DNA were then eluted by a high temperature wash. The cDNA present in this eluate (i.e. that which had been present in an IRD) was then PCR amplified using primers designed to the cDNA-specific linkers.

25µl M280 streptavidin coated magnetic beads (Dynal) were washed in phosphate buffered saline/BSA according to the manufacturer's instructions and then incubated for 15 mins at RT in 100µl TEN-S/P (0.1% SDS, 0.5mg/ml PVP in TEN buffer (TEN buffer: 10mM Tris.HCl, pH7.4, 1mM EDTA, 100mM NaCl)). This buffer was then replaced with 100µl TEN-S/P containing 10µg sonicated salmon sperm DNA and incubated for a further 15 mins. This buffer was then removed and the beads resuspended in 25µl TEN-S/P.

The duplexed DNA was made to 0.1% SDS, 0.5mg/ml PVP and then this solution was used to replace the bead solution. The DNA/bead mixture was incubated for 30 mins at RT with occasional mixing.

Washes were performed with 6 changes of 150µl TEN at RT, then 8 changes of 150µl 0.1x SSC at 68°C. The product was eluted by heating the beads to 90°C for 4 mins in 10µl dH_20 before immobilisation of the beads and recovery of the buffer/eluate DNA. The eluate was then passed through a further magnetic separation to remove all traces of beads.

2.14.6 Selection of the IRDs (EL-CSC)

The protocol follows that of HF-CSC until the duplexed DNA has been incubated with the beads.

Washes were performed with six changes of 150µl TEN at RT and then four changes of 150µl 0.1x SSC at 68°C. The beads were then resuspended in 20µl TEN and heated to 60°C before adding, in 1µl, 250ng capture oligonucleotide 732 and 250ng kinased capture oligonucleotide 735 (table 2.2). This was then left to cool over a period of 1 hour and washed in 4 changes of 100µl TEN. A 5 minute ligation reaction was then performed at 45°C in 20µl volume, using 10 units of Taq DNA ligase (Cambio), and then the reaction was stopped by adding 1µl 500mM EDTA. The sample was then washed with 3 changes of 100µl TEN at 45°C and 5 changes of 0.1x SSC at 65°C.

2.14.7 PCR amplification of the product

1µl of the eluate and controls were amplified in duplicate with 600ng of the cDNA linker primer 727 (HF-CSC) or 789 and 596 (EL-CSC) (table 2.2). The programme used was:

Denaturation: 93°C for 30 secs in the 1st cycle and 15 secs thereafter.

Annealing: 54°C for 30 secs

Extension: 72°C for 1½ mins (cycles 1-10), 2½ mins (cycles 11-20), 4 mins (cycles 21-29) then 8 mins (cycle 30)

Also 0.4µl of the beads (with YAC DNA attached) was amplified with 600ng of the YAC linker primer 477 (HF-CSC) (table 2.2), using the first 20 cycles of the PCR programme above.

The duplicate PCR reactions were pooled and then 8µl was visualised on a 2% agarose gel.

2.14.8 Cloning of the product DNA

A library was made from the product DNA by cloning the PCR product into pBS vector.

20µl of the product DNA was chloroform/isoamyl alcohol then ether extracted and ethanol precipitated. The pellet was resuspended in 5µl. The DNA was then heat shocked at 65°C for 5 mins and cooled to RT, to disentangle the molecules. The concentration was checked by running 1µl on a 2% agarose gel, and then the remainder was digested with *EcoRI* in a total of 10µl. The enzyme was heat killed at 70°C for 10 mins, and then the sample was ethanol precipitated. The pellet was resuspended in 10µl and subjected to heat shock as above.

20ng of the digested product DNA was ligated to 20ng pBS DNA (digested with *EcoRI* and phosphatased by Tony Brookes), in a total of 20µl at 16°C overnight. The ligation mix was then drop dialysed and 1µl was used for electro-transfomation of competent cells.

2.15 Extraction of RNA and Preparation of cDNA

2.15.1 Extraction of RNA

The extraction of RNA from tissue samples was performed using RNazol B™ (Biogenesis Ltd.). RNazol B promotes the formation of complexes of RNA with guanidinium and water molecules, and abolishes hydrophilic interactions of DNA and proteins. During the extraction step, the RNA remains in the aqueous phase, while DNA and proteins are removed from it.

Tissues from which the RNA was to be extracted were frozen as quickly as possible and then stored at -70°C. To extract the RNA, 1ml RNazol B was added to a small piece of the tissue (~1cm^3) in a plastic bijou, and the tissue was homogenised using an electric homogeniser. The homogeniser had been pre-sterilised by soaking in dH_20 containing 0.1% diethyl pyrocarbonate (DEPC) and then baking at 80°C for at least 1 hour. In between homogenising different tissue samples the homogeniser was rinsed first in

Virkon, then in dH_20 and finally in ethanol. After homogenising the sample was kept on ice. The following steps were carried out in a fume hood.

The RNazol B solution was transferred to a screw-cap eppendorf tube containing 100µl chloroform. The tube was inverted repeatedly for 15 secs and then placed on ice for 15 mins before centrifugation at 11,000 rpm for 15 mins at 4°C. The upper, aqueous phase was then transferred to a second screw-cap tube, 1 volume of isopropanol was added, and the tube left on ice for 15 mins. The tube was then centrifuged as before, and the supernatant discarded. The pellet was washed twice in 75% ethanol, and the open top covered with a small piece of Parafilm (American National Can Ltd.) which had been pierced many times with a needle. The pellet was then vacuum dried for not longer than 10 mins. The pellet was dissolved in 100µl of DEPC-treated water by agitating it at 4°C overnight.

The concentration of the RNA was calculated by reading the absorbance of a 1 in 200 dilution in DEPC-treated water at 260nm and 280nm. The concentration of the RNA (in mg/ml) was obtained by multiplying the Abs_{260} of the diluted sample by 10. The ratio of the Abs_{260} to the Abs_{280} provides a measure of the quality of the RNA. All values obtained were within the expected range (1.6 to 2.0). RNA samples split into aliquots of 10µl and stored at -70°C.

2.15.2 DNase treatment of RNA

For RNA samples to be used as a template for RT-PCR, it was necessary to pre-treat the RNA with DNase. This ensures that any RT-PCR products are genuinely derived from RNA template and not DNA. The efficiency of the DNase treatment can be checked by including a minus reverse transcriptase control during first strand cDNA synthesis (see section 2.15.3).

4µg RNA was treated with 10 units/µg DNase I (Boehringer Mannheim) and 10 units/µg RNase Inhibitor (Boehringer Mannheim) in a total of 40µl 1x

Transcription buffer (Ambion) at 37°C for 30 mins. It was then extracted with an equal volume of a 1:1 ratio of phenol:chloroform, and precipitated with 1/10th volume of 5M ammonium acetate and 3 volumes 100% ethanol at -70°C for 30 mins. The tube was then centrifuged for 15 mins at 11,000 rpm at 4°C. The pellet was vacuum dried and resuspended in 8µl DEPC-treated dH_20. A recovery of 50% of the RNA was assumed.

2.15.3 First strand cDNA synthesis

The volume of 2µg of DNase-treated RNA was adjusted to 8µl, and then it was heated to 65°C for 10 mins, and cooled on ice. To this 8µl was added the components of the First strand cDNA synthesis kit (Pharmacia Biotech), that is 5µl bulk strand reaction mix, 1µl DTT and 1µl $pd(N)_6$ primer. The bulk strand reaction mix contains Moloney Murine Leukaemia Virus reverse transcriptase which catalyses the polymerisation of dGTP, dCTP, dATP and dTTP into a single cDNA strand. The reaction was left to proceed at 37°C for 1 hour, after which the cDNA was used directly in a PCR reaction (5µl per reaction).

Table 2.2 Sequences of oligonucleotides used for PCR amplification and DNA sequencing

Key:

spl	splinkerette	LS	lower strand	LHS	left hand side	for	forward
vec	vectorette	ext'l	external	RHS	right hand side	rev	reverse
US	upper strand	int'l	internal	CL	catch linker		

NAME	FUNCTION	SEQUENCE 5'→ 3'	COMMENTS
Splinkerette and Vectorette PCR			
931	spl US	CGA ATC GTA ACC GTT CGT ACG AGA ATC GCT GTC CTC TCC AAC GAG CCA AGG	
930	spl LS	GAT CCC TTG GCT CGT TTT TTT TTG CAA AAA	
53	spl / vec ext'l primer	CGA ATC GTA ACC GTT CGT ACG AGA A	
52	spl / vec int'l primer	TCG TAC GAG AAT CGC TGT CCT CTC C	
861	CFTR ext'l primer	CCT GTC TCC TGG ACA GAA AC	obtained from Julia Dorin
860	CFTR int'l primer	CAA ATG AAT GGC ATC GAA	obtained from Julia Dorin
50	WT1 ext'l primer	CCA AGG GCC GTG AGG ATA CGC GAA G	obtained from Kathy Williamson
51	WT1 int'l primer	GCA CGC AGG CGC TGG CCC CCG ACA T	obtained from Kathy Williamson

Plasmid vector primers			
291	pBS primer	CAG GAA ACA GCT ATG AC	reverse primer (Stratagene)
292	pBS primer	GTA AAA CGA CGG CCA GT	forward primer (Stratagene)
YAC end-cloning			
227	pYAC4 LHS	CCC GTC CTG TGG ATC AAT TC	reverse of 372 from Arveiler and Porteous, 1991
373	pYAC4 LHS	GCT CCT TCC TTC GTT CTT CC	from Arveiler and Porteous, 1991
225	pYAC4 RHS	GCC AAG TTG GTT TAA GGC GC	reverse of 374 from Arveiler and Porteous, 1991
556	pYAC4 RHS	GCC CGA TCT CAA GAT TAC G	from Arveiler and Porteous, 1991
464	pYAC4 RHS	CCA CCA TAC CCA CGC CGA AAC AA	
Coincident Sequence Cloning			
477	YAC CL US & primer	CCG AAT TCT AGA GTC GAC C	+/- biotin at 5' end
479	YAC CL LS	GAT CGG TCG ACT CTA GAA TTC GG	
727	cDNA CL US & primer	GCG AAT TCT AGA CTG CAG G	
731	cDNA CL LS	AAT TCC TGC AGT CTA GAA TTC GC	
732	EL capture oligo 1	GGA CGG GTC GAC ACG CGA GGA CCG AAT TCT AGA GTC GAC C	partly complement of 479

735	EL capture oligo 2	GAT CGG TCG ACT CTA GAA TTC ACC CGT GCT ACC GGA ACG	partly complement of 477
596	EL product primer 1	GGA CGG GTC GAC ACG CGA GG	identical to part of 732
789	EL product primer 2	CGT TCC GGT AGC ACG GG	complement of part of 735
Analysis of the alpha tubulin gene			
711	PCR and/or sequencing primer	CTG TCA CAG GTG GGA GGG GAT TAA AGA GAT ATA AAC AAC TA	upstream of D0485 α tubulin sequence
712		GTC GAC TTT CAA TTT TCA AGT CTC ATT TGT TCA GCA GTA CAT TTC A	downstream of D0485 α tubulin sequence
496		GGG AGG AGA TGA CTC CTT CAA CAC CTT CAG	specific to D0485 α tubulin
497		ATG GCT GTG GTG TTG CTC AGC ATG CAC TCT	specific to D0485 α tubulin
150		AGA TGC CAA GTG ACA AGA CCA	
151		TGC AGG TCA ACA TTC AGG GCT	
152		TTA CTT ACC TCA ACT CTT AGC	
153		GCT CAT CAC AGG CAA GGA AGA	
154		TTT CAG GGC TTC TTG GTT TTC	
155		CGC CCA ACC TAC ACT AAC CTT	

156	*contd.*	CTC CCA CAT CCA CTT CCC TCT	
157		CAC CAA ATG AAA ATC TAA CCA	
158		TCA ACA GAA TCC ACA CCA ACC	
159		GGA AGA AAG AGA GGA ATA CTA	
309		TTC CCA GTG CAT GTA AAA GTC	
312		AGG CAC AAT CAG AGT CGT CCA	
313		TTC TGG TTA GAT TTT CAT TTG	
417		GGT TGG AGA CGC ATT CAC GCA	
467		ACT TAG GTG AAA ATG CTA TCG	
468		GCG TGA ATG CGT CTC CAA CCA	
469		AGA CAA TGA GGC CAT CTG TGA	
471		ACG CTT GGC ATT CAT CAG GTC	
709		TGG TGA TTG CTG AAG GTT GGA	
710		CAG TAA AGG CTC AAC ATA ATC	
714		CGG TAG GTG CCA GTG CAA ACT TCA TC	
715		GAC CAA GCG TAG CAT CCA GTT TGT AG	
1000	3' UTR probe	GGC TGC CCT TGA GAA GGA TTA	

Analysis of the other CSC products			
482	Family 1 US	AGG AGA AGC TCG TGG CTG CTG AGT T	
483	Family 1 LS	TTG TGA GTA GAA AGC ATT GGA GAG A	
484	Family 2 US	CCC AAG AAT GAG CAG AAG AAC AGG T	
485	Family 2 LS	ATT CGG TGA TGC CTC TGT GTC TGT C	
486	Family 3 US	CAT ATC TCA CAT CCT CAG AAA GGC T	
487	Family 3 LS	CAC AGA TSC AAA AGC CCC AGC CAG T	S= C or G
488	Family 4 US	TCC AAA TGG CAT CTC CTA CCC TAT C	
489	Family 4 LS	GCA AYG CAA GGA CAT CTA ATA CAG C	Y= C or T
490	Family 5 US	CAA AAC CTG ATT CAA ACT TGT GTG T	
491	Family 5 LS	GCG GGG CAG GAG TGT GTA CTT CTT C	
492	Product 6 US	ACC TCT CAT TTT CCC CCA GAC ATT T	
493	Product 6 LS	GTC ACT GTC CAC GGC AGC CCA GCA T	
494	Product 7 US	TCA GGA AGA AGG AAA GTA ACA AAT G	
495	Product 7 LS	TCC ACA CAT ATC TTC TAT CAT CCC A	
RT-PCR positive control			
for	GAPDH exon3	TTC ATT GAC CTC AAC TAC AT	obtained from Elena Farini
rev	GAPDH exon 8	GTG GCA GTG ATG GCA TGG AC	obtained from Elena Farini

CHAPTER 3

SPLINKERETTE PCR AND PHYSICAL MAPPING OF YACS

3.1 Introduction

3.1.1 PCR walking

PCR is widely used for the exponential amplification of sequences between two known oligonucleotide primers. The requirement for primers at both ends of the molecule to permit amplification is also the technique's major drawback, since this implies that some knowledge of DNA sequence at both ends of the desired product is necessary. In most instances this limitation further implies a need for laborious cloning and sequencing steps before the region can be amplified. Many methods have therefore been developed which circumvent the need for sequence information at both ends of the molecule, and enable the amplification of a product which extends outwards from a region of known sequence. This is termed 'PCR walking'.

The majority of published methods enabling PCR walking are based upon one of three themes. Those themes are i) randomly primed PCR, ii) inverse PCR and iii) linker-adapter PCR.

Randomly primed PCR (Fig. 3.1) relies on interactions between imperfectly base-paired primer and template to initiate priming. In order for a primer to function, it does not need perfect match with the template along its whole length. Indeed, under sufficiently low stringency conditions, successful priming can result when the primer has only partial 3' end homology with the template, with a perfect match of only two bases at the 3' end (Parker *et al*, 1991). PCR walking can thus be achieved employing a single sequence specific primer and random 'walking' primers (Parker *et al*, 1991, Dominguez and López-Larrea, 1994). Nested PCR using a sequence specific primer internal to the first, in combination with a walking primer, will increase the

specificity of the reaction. An alternative is to use the sequence specific primer alone, which at low annealing temperatures will initiate priming from non-specific sites where there is some chance homology at the 3' end (Parks *et al*, 1991). Thermal Assymetric Interlaced (TAIL) PCR (Liu and Whittier, 1995) overcomes some of the problems of background products encountered in randomly primed PCR, by alternating high and low stringency cycles. High stringency cycles allow priming from a long sequence specific primer only. Low stringency cycles also permit priming from a short arbitrary degenerate primer. The combination of the two, followed by two rounds of nested PCR will decrease the amount of product primed solely by the arbitrary primer.

The technique of Inverse PCR (Ochman *et al*, 1988) (Fig. 3.2) utilises two primers from the region of known sequence, which face in opposite directions. The DNA is digested with a restriction enzyme which has no site between the primers, and then it is re-ligated under conditions which favour the formation of monomeric circles. The two sequence specific primers may then be used to amplify around the circle of DNA between them.

The various methods of Linker-adapter PCR (Fig. 3.3) are a popular choice for PCR walking. Linker-adapter PCR involves amplification between a sequence specific primer and a primer complementary to a downstream synthetic linker sequence which has been ligated on to restriction enzyme digested DNA. In 'oligo-cassette mediated PCR' (Rosenthal and Jones, 1990), linkers are ligated onto the ends of digested DNA, then sequence specificity is provided by linear amplification from a biotinylated sequence specific primer. This product may be isolated by attachment to streptavidin-coated magnetic beads. There can then follow an exponential PCR reaction on the isolated sequences between the same sequence specific primer and the linker primer. A similar method, in which linkers are ligated to genomic DNA *after* linear amplification from the sequence specific primers is termed

Ligation Mediated PCR (Mueller and Wold, 1989). This procedure was developed as a new method for in-vivo footprinting of DNA sequences.

Several linker-adapter PCR methods involve a specially designed linker, such that extension from the linker primer cannot take place without prior extension from the sequence specific primer. This is usually due to the absence of a sequence complementary to the linker primer at the beginning of the reaction. The sequence to which the linker primer can anneal must be synthesised by extension from the sequence specific primer. In this way, artefactual linker-linker products are not produced. In the simplest case, this may involve a linker which has not been kinased at the 5' end of the bottom strand, such that there is no covalent bond formed between the linker sequence complementary to the primer and the genomic DNA (Kalman *et al*, 1990). Extension from the linker primer will therefore terminate at the end of the linker until that strand has been generated by extension from the sequence specific primer.

Linkers in which the primer portion of the linker is single stranded will prevent extension from the linker primer (top strand) in the first cycle (Roux and Dhanarajan, 1990). This however provides opportunity for extension from the 3' end of the bottom strand to generate a sequence complementary to the primer (Upcroft and Healey, 1993). This may be prevented by the presence of a mismatched tail (Roux and Dhanarajan, 1990) or an amine group (Siebert *et al*, 1995) on the bottom strand.

Figure 3.1 Randomly Primed PCR (adapted from Parker *et al*, 1991)

The targeted and internal primers are complementary to a known sequence. The walking primers are of arbitrary sequence. The walking primers will hybridise to multiple sites on both DNA strands, but a targeted PCR product will only be produced when a walking primer anneals at a suitable distance from the targeted primer on the complementary strand. Nested PCR using the internal primer and a walking primer will increase the specificity of the reaction.

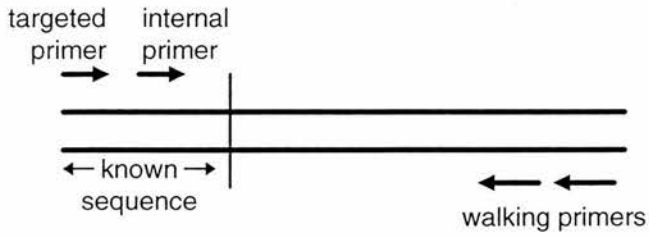


Figure 3.2 Inverse PCR (adapted from Ochman *et al*, 1988)

The core region (known sequence) is depicted as a thick straight line. The filled and open boxes represent upstream and downstream flanking regions respectively.

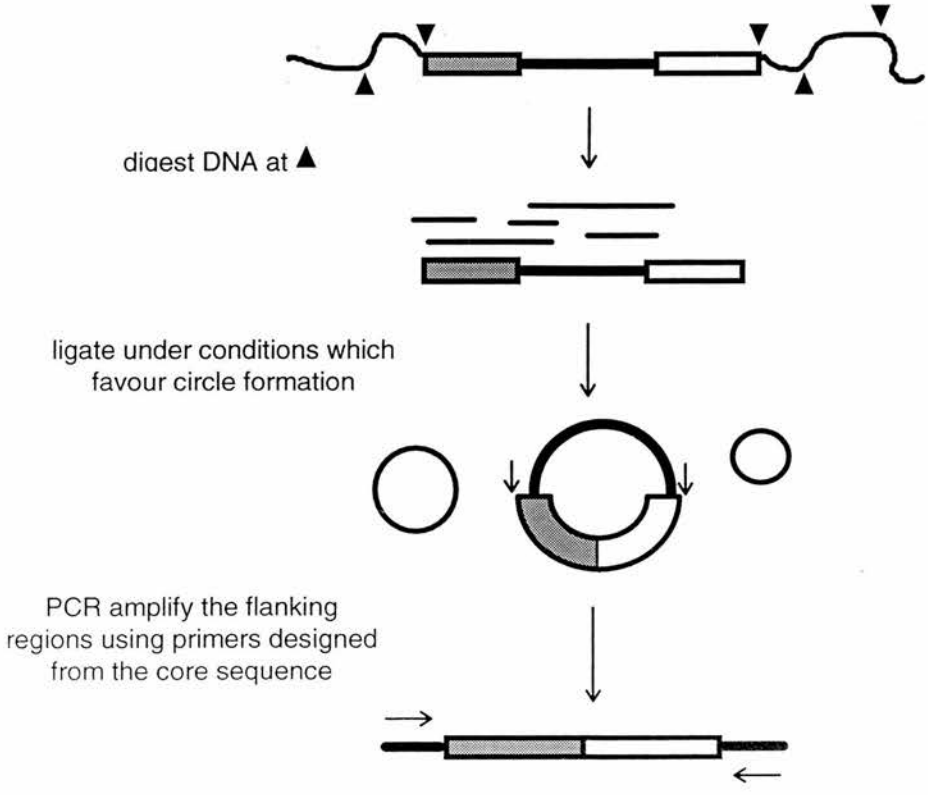
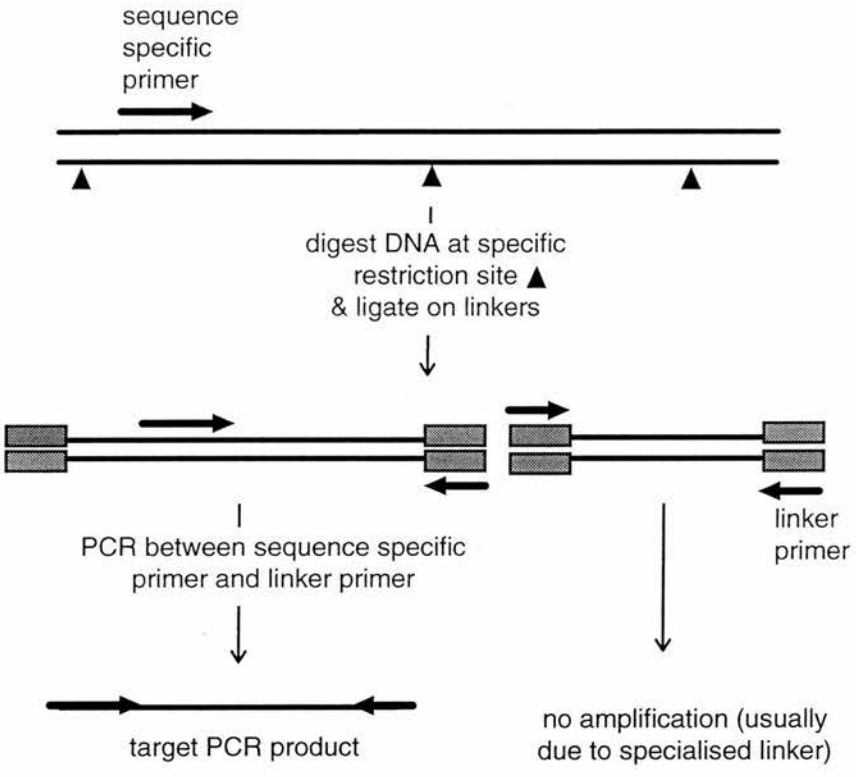


Figure 3.3 Linker-Adapter PCR

Different methods employ different strategies to avoid amplification between two linker primers (not shown), for instance mismatched linker strands (Roux and Dhanajaran, 1990), an amine group on one linker strand (Siebert *et al*, 1995), unligated linkers (Kalman *et al*, 1990) or preselection of material amplified by the sequence specific primer by binding to magnetic beads (Rosenthal and Jones, 1990).



3.1.2 Vectorettes

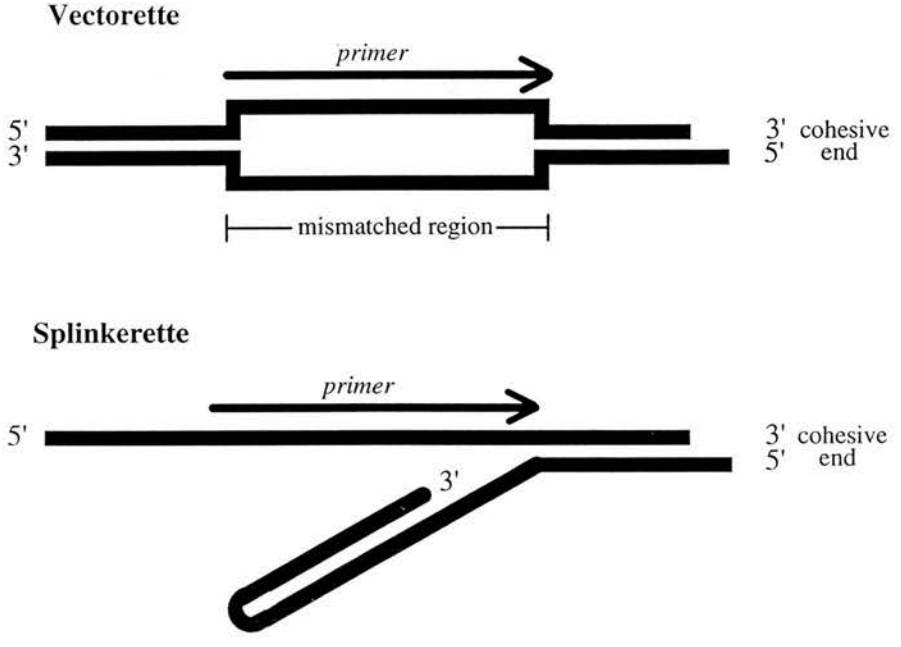
The vectorette™ is an example of a specially designed linker (Arnold and Hodgson, 1991 and Riley *et al*, 1990). It comprises a double stranded sequence of ~50bp, with one cohesive end rendering it suitable for ligation to restriction enzyme digested DNA (Fig. 3.4). The associated vectorette primer and nested primer are of the same sequence as part of the top strand of the linker. However, sequences complementary to these primers are not present in the linker, since there is a 'bubble' of mismatched bases between the top and bottom strands. The 3' end of the vectorette primer is situated within the mismatched region. Extension from a sequence specific primer however will copy the top strand into its complement, and hence form a priming site for the vectorette primer. Exponential amplification may then ensue between the sequence specific primer and the vectorette primer.

The design of the vectorette linker is such that it should prevent amplification of molecules possessing vectorettes at both ends. The specificity of the vectorette PCR reaction is however not absolute, especially for amplification of longer PCR products (>2kb) from uncloned genomic DNA. Illegitimate products resulting from non-specific annealing of either primer may be minimised by the use of a sufficiently high annealing temperature. However, background products may also result from a phenomenon we term 'end-repair' priming, which is intrinsic to the design of vectorettes (Devon *et al*, 1995). End-repair priming is caused by the presence of the free cohesive ends of unligated vectorettes and genomic DNA. The cohesive ends possess a 5' overhang which may be filled in during the first extension step of the PCR reaction. Following denaturation, these filled-in ends are able to anneal to each other, and the 4bp overlap thus produced is sufficiently stable to be able to initiate priming (Sommer and Tautz, 1989). This enables extension from the genomic DNA across the vectorette, thus producing a sequence complementary to the vectorette primer. It also enables extension from the vectorette along the genomic DNA.

3.1.3 Splinkerettes

The splinkerette, an alternative to the vectorette, was developed in response to the prevalent end-repair priming in vectorette PCR. The splinkerette is a specialised linker sequence, used exactly as the vectorette, but its design is such that end-repair priming can not occur (Devon *et al*, 1995). Instead of a mismatched region between the top and bottom strands, the splinkerette incorporates a hairpin structure on the bottom strand (Fig 3.4). After the first denaturation step of the PCR reaction, the free 3' end of the bottom strand will flip back on itself to form the hairpin, and may then begin elongation further along the bottom strand. The elongation will continue until the end of the bottom strand is reached, thus forming a stable double stranded structure which is not available for end-repair priming. The bottom strand of the splinkerette is not kinased, and hence there is no covalent bond between it and the genomic DNA. This prevents elongation from the hairpin along the whole length of the genomic DNA to the splinkerette linker at the other end.

Figure 3.4 Schematic Diagram of the Vectorette and Splinkerette (from Devon *et al*, 1995)



3.2 A Comparison of Vectorettes and Splinkerettes

A practical comparison of the ability of vectorette and splinkerette PCR to execute efficient PCR walking on total human genomic DNA was undertaken. Model systems were chosen for target PCR products where primers and/or sequence information was available, and the size of the expected PCR product was known. Two target PCR products were chosen. The first was a fragment from exon 13 of the cystic fibrosis transmembrane regulator (CFTR) gene on chromosome 7. The distance between available primers and the next *Sau3AI* site, to which the linker could be ligated, was calculated from sequence information to be ~290bp. The second was a product of from the first intron of Wilm's Tumour 1 (WT1) gene on chromosome 11. The target product from the primers to the next *BamHI* site (also compatible with the linkers) was estimated to be ~2.3kb from a published restriction map of the gene (Tadokoro *et al*, 1992). For both products, two primers were available in the same orientation to enable nested PCR to be carried out, and an oligonucleotide (for CFTR) or a clone (for WT1) was available to be used as a probe to verify the specific nature of the PCR products.

To enable direct comparison, the splinkerette and vectorette were designed to be compatible with the same external and nested primers. Total human genomic DNA was digested with *Sau3AI* or *BamHI*. The splinkerette and vectorette used possessed a 4bp GATC overhang, thus rendering them suitable for ligation to *Sau3AI*, *BamHI*, *Bgl II* or *Bcl I* digested DNA. Splinkerettes or vectorettes were ligated on to the ends of the digested genomic DNA molecules. For the 290bp product, ligation was with a 3:1 molar ratio linker:insert. For the 2.3kb product, the molar ratio used was 15:1 linker:insert, as recommended for the vectorette by ICI. Vectorette ligations were performed for 4½hrs at RT as recommended by ICI. Splinkerette ligations were initially performed overnight at 4°C (used for the 290bp product PCR), but it was found that a superior result was produced with a

4½hr ligation at RT, as for the vectorette. These ligation conditions were therefore used for the 2.3kb target product.

3.2.1 The 290bp product

PCR amplification was performed on *Sau3AI* digested, linker ligated DNA, using the external CFTR primer, 861, plus the vectorette/ splinkerette primer. Nested PCR was then carried out using the internal CFTR primer, 860, plus the nested vectorette/splinkerette primer (see table 2.2 for primer sequences). The PCR programmes used were:

Primary PCR

Hot start for 5 mins at 90°C.

Denaturation: 93°C for 30 secs in first cycle and 15 secs thereafter.

Annealing: Touch down from 65-55°C for 1 min.

Extension: 72°C, 5 secs (cycles 1-10), 15 secs (11-20), 45 secs (21-30).

Nested PCR

No hot start

Denaturation: 93°C for 30 secs in the first cycle and 15 secs thereafter.

Annealing: 55°C for 1 min.

Extension: 72°C, 5 secs (cycles 1-10), 15 secs (11-20), 45 secs (21-30).

After the primary PCR, the spectrum of products was very similar for the vectorette and the splinkerette (Fig 3.5). Both produced a major band at 290bp plus other bands and a background smear. Upon reamplification with the nested primers however, this was reduced to a single band of the expected size for both linkers, with equal intensity between the vectorette and the splinkerette (Fig 3.5). When the nested primer, 860, was used as an oligonucleotide probe on to the primary PCR products, the major band hybridised strongly with equal intensity for the vectorette and splinkerette tracks (Fig 3.5).

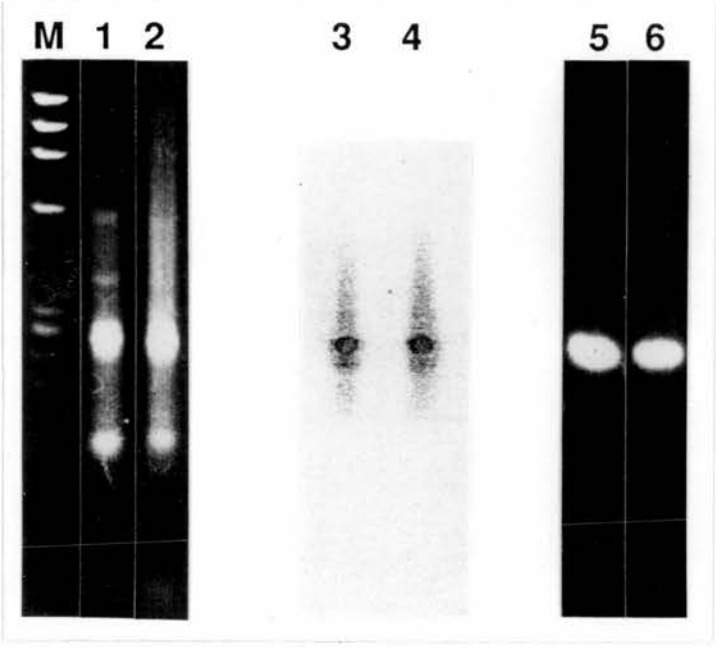


Figure 3.5 Amplification of a 290bp product by splinkerette or vectorette PCR
The marker used is 250ng ϕx *Hae III*
Tracks 1 and 2 show 8µl of the vectorette and splinkerette primary PCR product respectively (generated using the external CFTR primer (861) plus the vectorette/splinkerette primer), resolved on a 4% agarose gel.
Tracks 3 and 4 show the result of hybridisation of tracks 1 and 2 respectively with the internal CFTR oligonucleotide (860).
Tracks 5 and 6 show 10µl of the vectorette and splinkerette secondary PCR product respectively (generated using the internal CFTR primer (860) plus the nested vectorette/splinkerette primer), resolved on a 4% agarose gel.
It can be seen that vectorette PCR and splinkerette PCR perform equally under these conditions.

3.2.2 The 2.3kb product

Primary PCR amplification of linker ligated *BamHI* digested DNA was performed using the external WT1 primer, 50, plus the vectorette/ splinkerette primer. Nested PCR was then carried out using the internal WT1 primer, 51, plus the nested vectorette/ splinkerette primer (see table 2.2 for primer sequences). The PCR programmes used were:

Primary PCR

Hot start for 5 mins at 90°C.

Denaturation: 95°C for 30 secs in the first cycle and 15 secs thereafter.

Annealing: Touch down from 70-60°C for 30 secs.

Extension: 72°C, 2 mins (cycles 1-10), 4 mins (11-20) and 6 mins (21-30).

Nested PCR

Hot start

Denaturation: 95°C for 30 secs in the first cycle and 15 secs thereafter.

Annealing: Touch down from 71-61°C for 30 secs

Extension: 72°C, 2 mins (cycles 1-10), 4 mins (11-20), 6 mins (21-30).

No product was visible after primary PCR for either the vectorette or the splinkerette. However upon reamplification with the nested primers, a strong band of 2.3kb upon a faint background smear was visible for the splinkerette, whereas an intense smear only was produced by the vectorette (Fig. 3.6). These products were hybridised with a 1.1kb *Xho I* and *Not I* restriction fragment from the clone WTex1 (obtained from Andreas Schedl), which contained part of the target PCR product. The 2.3kb band in the splinkerette track hybridised strongly to the probe, verifying that it was the target PCR product. A band of the right size was also seen in the vectorette track, but its intensity was very low (Fig. 3.6).

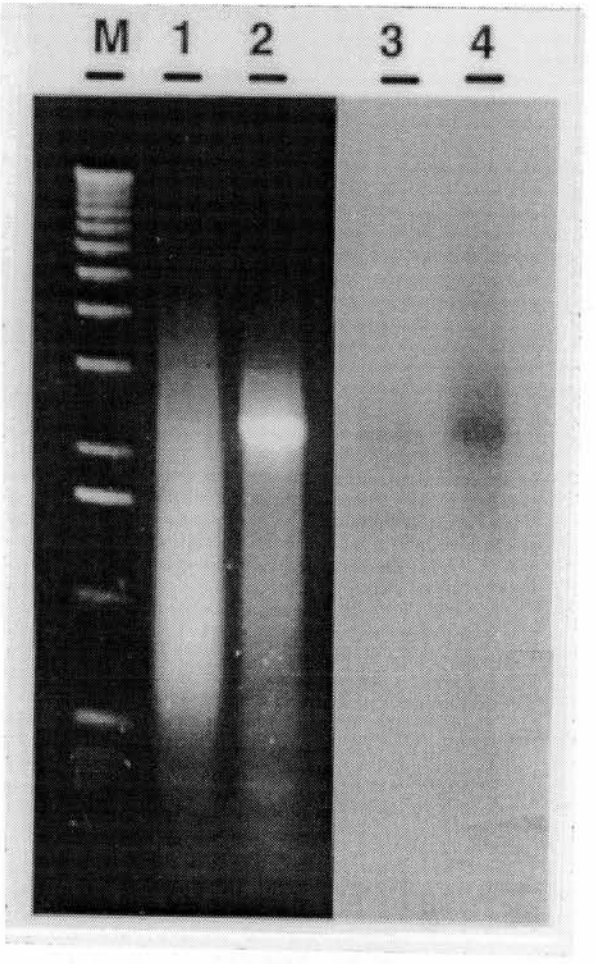


Figure 3.6 Amplification of a 2.3kb product by splinkerette or vectorette PCR (taken from Devon *et al*, 1995)

The marker used is 250ng of the 1kb ladder.

Tracks 1 and 2 show 10µl of the vectorette and splinkerette secondary PCR product respectively (generated using the internal WT1 primer (51) plus the nested vectorette/splinkerette primer), resolved on a 1% agarose gel.

Tracks 3 and 4 show the result of hybridisation of tracks 1 and 2 respectively with a 1.1kb probe derived from within the target PCR product.

It can be seen that the decreased non-specific priming associated with splinkerette PCR relative to vectorette PCR permits the generation of the 2.3kb product by splinkerette PCR only.

3.2.3 Discussion

Splinkerette PCR and vectorette PCR both performed equally in the amplification of a short, 290bp fragment from total human genomic DNA. Both enabled the production of the correct fragment (as verified by hybridisation) after one round of PCR, but other bands and a smear were also produced. Only after a second round of nested PCR was the correct band cleanly produced.

Splinkerette PCR was however significantly more effective than vectorette PCR in the amplification of a 2.3kb fragment. After two rounds of PCR the correct product was produced as a strong band by splinkerette PCR, but vectorette PCR resulted in a non-specific smear obscuring the real product.

The difference in competence between splinkerette and vectorette PCR can be attributed to the different structures of the linkers themselves. The hairpin structure of the splinkerette implies that it cannot partake in end-repair priming, which may be responsible for most of the background products seen in vectorette PCR. The differences are not caused by variation in primer efficiency, since the same primer and PCR conditions were used for the splinkerette and vectorette. Ligation conditions and genomic DNA were also identical for both.

The superior functioning of splinkerettes was seen only for the longer target PCR product. For the shorter target product, the vectorette and splinkerette performed equally. The reason for this may lie in the PCR bias towards the amplification of shorter molecules. When the target product was short, amplification of the specific product was favoured in the reaction. For a longer target product however, competition from smaller non-specific products is more significant. The low level of end-repair priming of splinkerettes reduced the competition from smaller non-specific products, allowing the genuine product to be made. Amplification of the target product

by vectorette PCR was however significantly encumbered by competition from small non-specific products.

3.3 Isolation of End Clones from YACs

3.3.1 Introduction

The isolation of the terminal sequences of recombinant DNA from YACs is an important step in the construction of a YAC contig. Hybridisation or PCR mapping of end-clone sequences onto other YACs provides a quick and easy way to score for overlap between them and will also determine which YACs extend the furthest at both ends of an emerging contig.

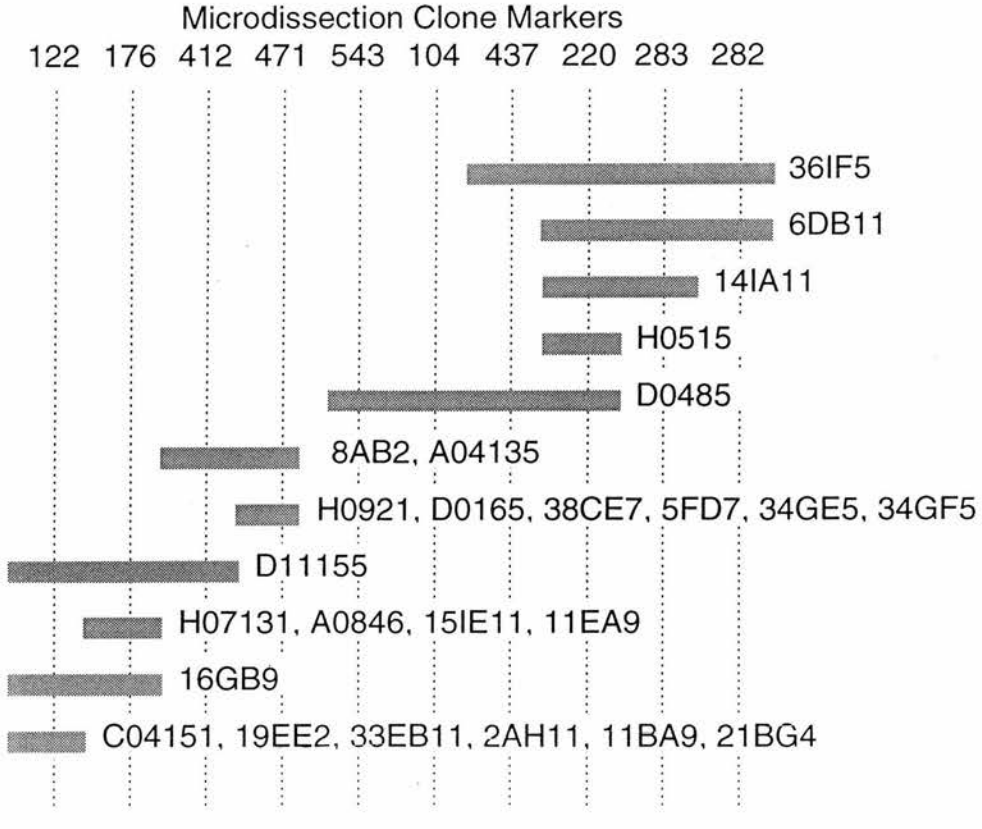
End clones can be isolated by subcloning of the YAC into cosmids or plasmid libraries and then selected by hybridisation with the YAC vector arms. The technique of 'junction trapping' (Patel *et al*, 1993) involves the isolation of end clones by subcloning the YAC into plasmids and the amplifying the end clones by performing PCR with the plasmid vector primer and the YAC vector primer.

Techniques which necessitate a cloning step can be laborious and time consuming, and for this reason PCR based techniques for the isolation of end clones are more commonly used. All the PCR based techniques make use of primers derived from the YAC vector arm to provide the end-specificity. The other primer may be derived from the Alu consensus sequence (Nelson *et al*, 1989), a linker such as a vectorette (Riley *et al*, 1990) or splinkerette (Devon *et al*, 1995), or even another primer from the YAC vector for inverse PCR (Ochman *et al*, 1988, Arveiler and Porteous, 1991).

3.3.2 Isolation of end clones from chromosome 11 YACs in the region of the t(1;11) breakpoint

A total of 23 YACs had been isolated by screening the ICI YAC library (Anand *et al*, 1990) and the ICRF Reference Library (Larin *et al*, 1991) with 5 repeat-free microdissection clones from the interval nearest the translocation breakpoint. These had been screened for positivity with the 11 microclones from this region, thus enabling the assembly of a first generation contig, which contained a single gap (Fig. 3.7). This contig however contained no information regarding overlap between individual YACs or the orientation of the contig.

Figure 3.7 First generation YAC contig. YACs (grey bars) had been screened by PCR for positivity with the microdissection clone markers (dotted lines). The contig contains a single gap, between YACs D0485 and 8AB2.



Splinkerette PCR was the method chosen for the isolation of end clones from the ICI YACs. A PCR-based method was chosen due to its speed and specificity. Splinkerette PCR had been proven to be at least as effective as vectorette PCR in PCR walking from total human genomic DNA (section 3.2) so it was assumed that it should also be competent in the isolation of YAC end fragments, perhaps succeeding in the amplification of longer products where vectorette PCR would have failed. (ICRF YAC end clones were isolated by vectorette PCR by John Brown.)

Approximately 0.5µg total yeast plus YAC DNA was double digested with *BamHI* and *Bgl II*. These enzymes have cohesive ends compatible with ligation to splinkerettes, and cut genomic DNA with a frequency suitable for the isolation of long end clones (*BamHI* average size fragment = 5.4kb; *Bgl II* average size fragment = 2.25kb (Drmanac *et al*, 1986)). Splinkerettes were then ligated onto the digested ends in a 4:1 ratio linker:insert. The ligated DNA was then PCR amplified with the splinkerette primer or nested primer and a primer derived from the YAC vector. YAC vector primers used were 227 or 373 ('left hand side'), and 225, 556 or 464 ('right hand side') (see table 2.2). Primers 225, 227, 556 and 373 were taken from or modified from Arveiler and Porteous, 1991. Primer 556 was used in some cases as the 'right hand side' primer, but it occasionally resulted in non-specific products from the yeast genome, so a new primer, 464, was designed from the pYAC4 sequence. A single round of PCR was generally sufficient to generate a single end-clone product free from background smear.

A standard PCR programme was used in all cases. This was:
Hot start for 5 mins at 90°C.
Denaturation: 94°C for 30 secs in the first cycle and 15 secs thereafter.
Annealing: Touch down from 68-58°C for 30 secs.
Extension: 72°C, 3 mins (cycles 1-10), 5 mins (cycles 11-20), 8 mins (cycles 21-30).

A total of 24 end clones were isolated, from 38 attempted (Fig. 3.8a). The average size of the fragments was 1.7kb. Table 3.1 lists all the ICI YAC end clones isolated, with their sizes. The bands were excised from a gel and used as hybridisation probes onto a panel of all the ICI YACs digested with *EcoRI* (Fig. 3.8b). Included on the filter was a track of total human genomic DNA and a track of total yeast genomic DNA, both digested with *EcoRI*.

Table 3.1 End Clones Isolated from ICI YACs

† = isolated by Euan Slorach
* = isolated by Sheila Christie
∞ = isolated by Kathy Evans
ND = not determined

YAC NAME	LHS/RHS FRAGMENT	pYAC4 PRIMER	LENGTH OF END CLONE (kb)	HUMAN HIGH COPY REPEAT
2AH11	L	227	1.0	-
5FD7	L	227	3.0	-
8AB2	L†	227	2.1	+
	R	556	1.2	-
11BA9	L†	227	1.4	+
11EA9	L†	227	1.1	-
	R	556	0.7	-
15IE11	R	225	2.1	ND
16GB9	L	227	0.6	+
	R	464	3.0	-
19EE2	L	227	1.4	+
21BG4	L	227	1.4	+
33EB11	L	227	1.1	-
	R	464	3.5	+
34GE5	L*	227	2.8	+
	R*	464	1.2	-
34GF5	L*	227	2.8	+
	R*	464	1.2	-
35IA2	L∞	227	4.0	+
36IF5	L†	227	0.8	-
	R	464	3.6	+
37GE11	L∞	227	0.6	-
37GE12	L∞	227	0.6	-
38CE7	R*	464	0.5	-

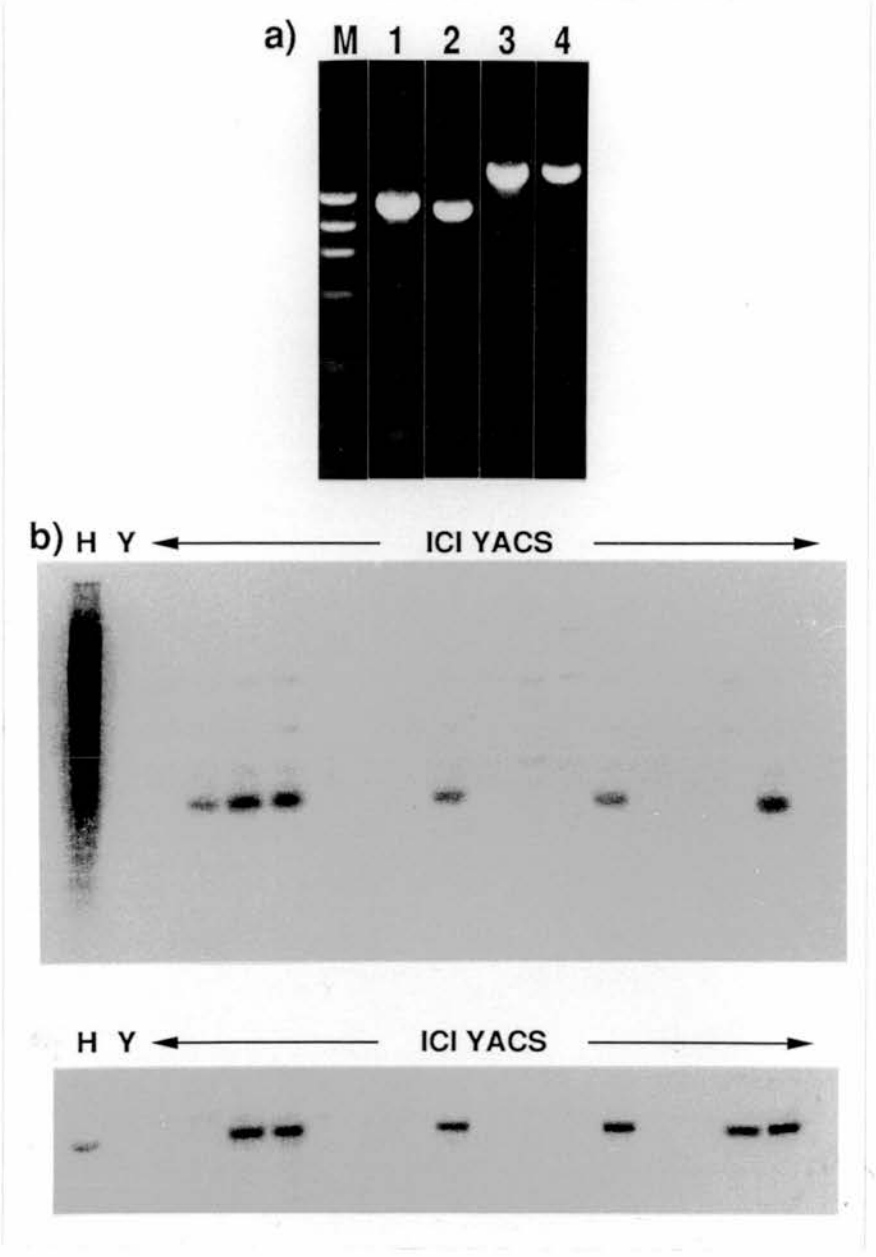


Figure 3.8 End Clones isolated from the ICI YACs

A) Tracks 1 to 4 show end clones from the left hand end of YACs 33EB11, 2AH11, 21BG4 and 19EE2 respectively. 10µl of the PCR product generated by splinkerette PCR using pYAC4 primer 227 is resolved on a 1% agarose gel. The marker used is 250ng ϕx Hae III. B) Hybridisation of left hand end clones from YACs 21BG4 (upper) and 2AH11 (lower) onto a panel containing *EcoRI* digests of the ICI YACs, total human genomic DNA (H) and total yeast genomic DNA (Y). The end clone from YAC 21BG4 contains a high copy human repeat and gives a signal on YACs 11EA9, 16GB9, 21BG4, 11BA9, 19EE2 and 15IE11 (left to right). The end clone from YAC 2AH11 is a single copy sequence and gives a signal on YACs 16GB9, 21BG4, 11BA9, 19EE2, 2AH11 and 15IE11 (left to right). In both cases there is insufficient signal in the yeast track to be visible.

3.3.3 Discussion

A single round of non-optimised splinkerette PCR cleanly generated an end clone in 63% of cases. Where failures occurred, it was usually because the PCR reaction resulted in multiple bands rather than a single one. The non-specific bands were derived from the yeast genome, since upon hybridisation to a panel of total yeast plus YAC DNAs, a band was seen in every track. The production of non-specific bands could have been reduced by using pure YAC DNA as the template for PCR, by optimisation of the PCR conditions and by implementation of a second round of PCR using nested primers. In some cases, the distance between the pYAC4 primer and the nearest *BamHI* or *Bgl II* site may have been too great for standard PCR. Long PCR conditions could have been used, or the YAC DNA could have been digested with enzymes other than *BamHI* and *Bgl II*. *Bcl I* and *Sau3AI* also create cohesive ends compatible with the existing splinkerette. The average size of *Sau3AI* digested fragments is small (325bp; Drmanac *et al*, 1986) so use of this enzyme would create a high probability that end clones could be isolated, but they would generally be too small to be useful after removal of the YAC vector DNA.

The average size of the end clones isolated was 1.7kb, which is a useful size for hybridisation probes. The pattern of signal seen upon hybridisation to a panel of the YACs provided some evidence that the PCR product probe was a real end clone, if it hybridised to the YAC from which it was derived, plus a discrete set of the other YACs. The final proof that the products were genuine end clones could have been derived by sequencing of the YAC vector junction with the recombinant DNA insert. This however was deemed not necessary for the end clones isolated, since only those in which a single strong band was produced in the PCR reaction were used, and no discrepancies were found when mapping them onto the panel of YACs.

The track of human genomic DNA was used to assess the repetitive nature of the probe. A single band of the same size as that on the YAC panel indicated a single copy probe, whereas probes which gave a smear on human genomic DNA were of a repetitive nature. High copy human repeats (probably Alu sequences) were present in 43% of probes tested. The Alu element is approximately 300bp long, one sixth of the average end clone length. This implies that Alu repeats constitute at least 8% of the total length of the genomic DNA in the end clones, which is close to the figure quoted for the whole genome (5%) (Shen *et al*, 1991).

The results generated from hybridisation of the end clones enabled scoring of overlap between YACs. It also provided data suggestive of chimaerism in YACs 36IF5, 33EB11 and 38CE7. This chimaerism was further investigated by FISH analysis of catch linkered YAC material (Shibasaki *et al*, 1995) (by Yoshiro Shibasaki and Diane Lawson). End clone data also identified clonal YACs, for which end clone PCR products were of the same size and gave identical mapping results. Three sets of clonal YACs were discovered. These were a group of three (19EE2, 11BA9 and 21BG4; 11BA9 only shown on Fig. 3.9), and two groups of two (37GE11 and 37GE12, 34GE5 and 34GF5; 37GE11 and 34GE5 shown on Fig. 3.9).

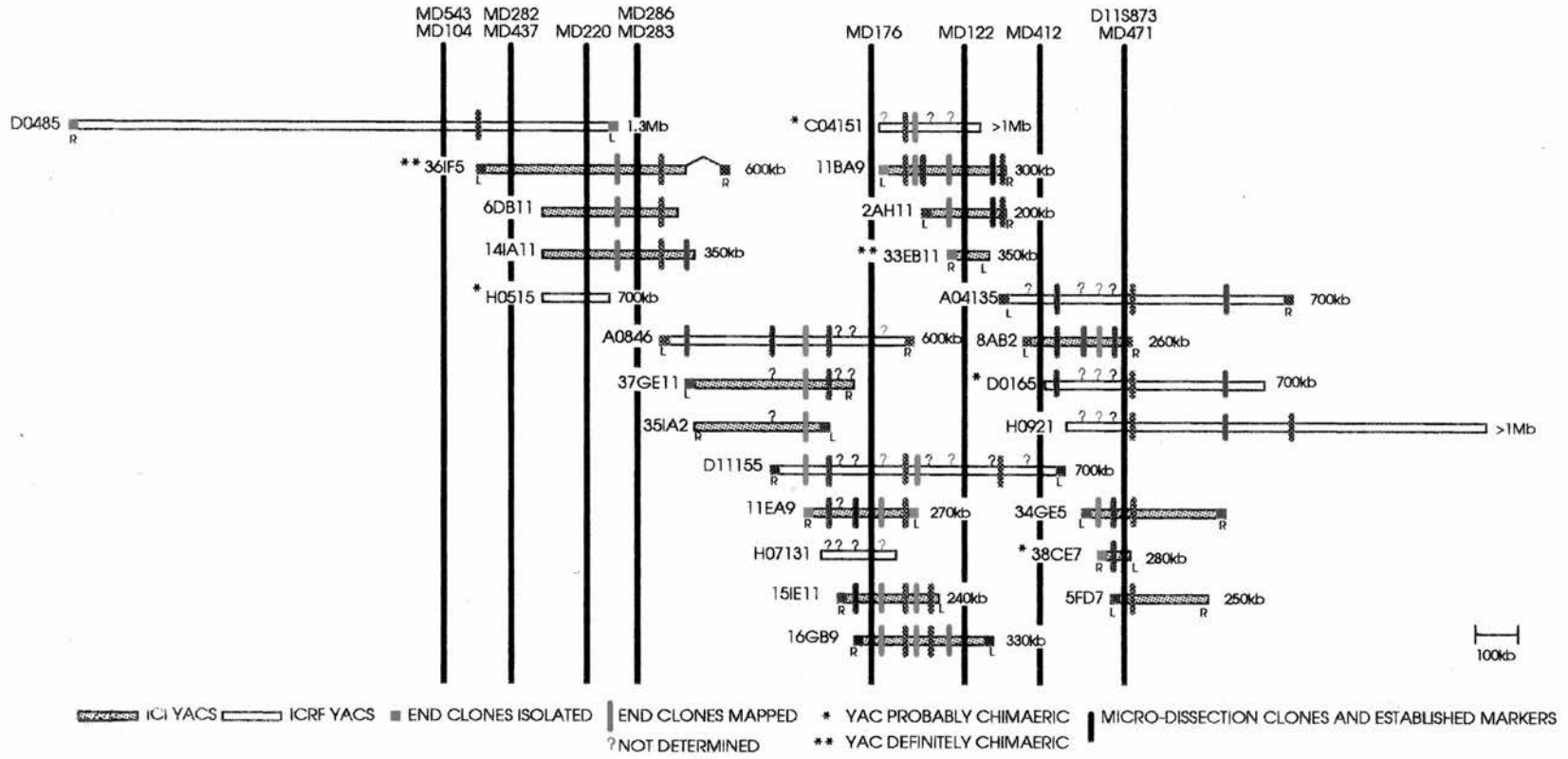
Cumulated data from mapping of these ICI YAC end clones, together with those isolated from the ICRF YACs, enable the final contig map to be drawn (Fig. 3.9). The order of microdissection clones 176 and 122 on the first generation contig map was reversed. In addition YAC 6DB11 was found not to be positive for MD282, and the position of the latter was moved to that of MD437. The gap in the contig which had been present before the addition of end clone data had been removed by the hybridisation of the left hand end clone from YAC A0846, and confirmed by the isolation of further YACs and end clones (37GE11, 37GE12 and 35IA2) (by Kathy Evans). The extent of overlap between YACs could not be judged by this preliminary end clone

mapping, but it could be estimated by combining the end clone data with results from pulsed field electrophoresis sizing of the YACs (by John Maule) and scoring of microdissection clone markers. The total length of the contig was approximately 3Mb.

Study of published linkage maps of the human genome (Gyapay *et al*, 1994) suggested that some genetic markers should lie within the contig. The position of the YAC contig on chromosome 11 was initially defined as between the gene tyrosinase (TYR) and the microsatellite marker D11S388, a distance estimated as 7.2cM. Généthon markers which lie in between these two markers are D11S931, D11S1342, D11S1358, D11S1332 and D11S873, all of which are microsatellites. The first three and the latter two fall into two groups among which the recombination rate is zero, and there is a genetic distance of 0.1cM between the two groups. PCR mapping on YACs from the contig enabled placement of these genetic markers on the contig (see Fig. 3.9). D11S931, D11S1342, D11S1358 and D11S1332 were located within YAC D0485, and D11S873 was within YAC A04135 (and others).

FISH analysis was performed (by Yoshiro Shibasaki and Diane Lawson) using catch-linkered YAC material and D11S931/MD104 positive cosmids isolated from ICRF chromosome 11 cosmid library (obtained from H. Lehrach). This established the orientation of the contig (as in Fig. 3.9) and also that YAC D0485 spanned the translocation breakpoint. The markers on this YAC were all located below the translocation breakpoint except D11S931, which lay an estimated 100kb above the breakpoint towards the centromere. Finer mapping of YAC D0485 then began with the subcloning of the YAC into cosmids, to establish more accurately the position of the translocation breakpoint and enable further analysis of the mechanism of the translocation. These further mapping efforts will be discussed briefly in Chapter 7.

Figure 3.9 A 3Mb contig of ICI and ICRF YACs on chromsome 11q14-21 (approximately to scale)



3.4 Long Range Restriction Site Mapping of YACs by Pulsed Field Gel Electrophoresis (PFGE)

Once the YAC contig had been constructed with end clone isolation and mapping, it was necessary to build a long range restriction map of the region, for finer mapping and to establish the position of genetic markers, and partly to look for restriction enzymes sites characteristic of CpG islands. I was responsible for the construction of the restriction map of two of the ICI YACs, 19EE2 and 8AB2, and was involved in the mapping of D0485 (see Chapters 5 and 6).

3.4.1 Restriction enzyme digestion and PFGE of YACs

Agarose plugs containing the YACs 19EE2 and 8AB2 (made by John Maule) were digested singly with the enzymes *Not I*, *BssH II*, *Sfi I*, *Sal I*, *Mlu I* and *Eag I* (the latter for 8AB2 only). The digest patterns were then resolved on a 0.5% pulsed field gel by running at 180V for 30 hrs with a pulse time of 50 secs. Markers used were 3B3, YP148 and concatemers of the phage λ genome, which give bands spaced every 48kb. The gel was Southern blotted.

Enzymes which did not cut within the YAC resulted in a single YAC derived band on the gel, which could be used to estimate the total size of the YAC. 19EE2 was estimated at 280kb and 8AB2 was 450kb. The total size of the YAC derived bands in every track should add up to this figure. If the total size of the YAC bands exceeded the total length of the YAC, then it was likely that the digestion had been incomplete.

3.4.2 Identification of the fragments derived from the YAC ends

Probes derived from the YAC vector arms were used to identify those fragments of human DNA at the ends of the YAC. The enzymes chosen for digestion do not cut within the vector sequence, to ensure that the nearest

enzyme sites were within the insert of human DNA, and that a single band would result from hybridisation with the vector arm probes.

2μg of pBR322 DNA was digested with the restriction enzymes *BamHI* and *Pvu II*. The digested DNA was then run on a 0.8% LMP agarose gel. Digestion with these two enzymes gave two bands, which were excised from the gel to be used as probes. The upper band (~2.8kb) was the left arm probe and the lower band (~1.7kb) was the right arm probe.

Sizes of the fragments hybridising to the left and right vector arms respectively are shown in Table 3.2.

3.4.3 Identification of other YAC derived fragments

Hybridisation with human Cot1 DNA was used to identify all the bands containing human DNA. Human Cot1 DNA is largely comprised of Alu repetitive elements, which are human specific and thus do not hybridise to yeast sequences. Alu elements are sufficiently frequent that it is highly likely that all digestion bands of a size resolved by PFGE (i.e. >10kb) will hybridise to human Cot1 DNA.

By subtraction it could be calculated which bands were present in the YAC but not derived from either end. In both YACs, for all the enzymes used, this resulted in a maximum of one extra band, which could therefore be placed between the two end fragments. Table 3.2 shows the sizes of bands hybridising to Cot1 DNA but not derived from either end of the YAC.

Table 3.2 Sizes of bands (kb) produced by hybridisation to restriction digested YACs 19EE2 and 8AB2.

Restriction Enzyme	Left vector arm	Right vector arm	Additional bands (Cot1 positive)
19EE2			
Not I	280	280	-
BssH II	280	280	-
Sfi I	50	150	80
Sal I	130	150	-
Mlu I	190	90	-
8AB2			
Not I	450	450	-
BssH II	160	160	130
Sfi I	220	230	-
Sal I	180	50	220
Mlu I	180	120	150
Eag I	120	250	80

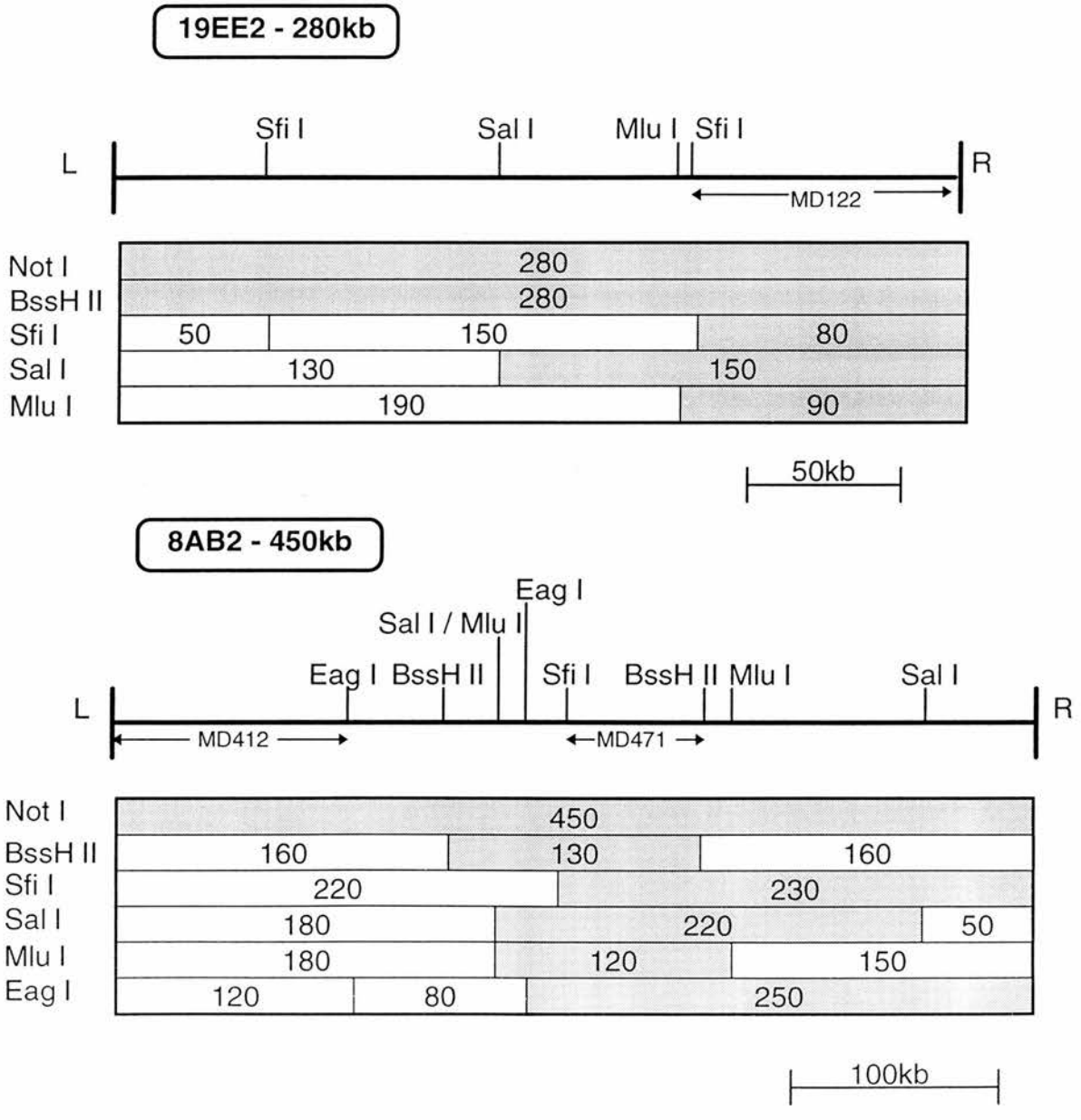
3.4.4 Placement of microdissection clone markers on the map

YAC 19EE2 is positive for one microdissection clone, MD122. Using the insert of this clone as a probe gave a pattern of hybridisation identical to that produced with the right hand end probe. This allowed placement of MD122 within the smallest fragment defined by the right vector arm probe, that is 80kb between the end of the YAC and the nearest *Sfi I* site (Fig. 3.10).

YAC 8AB2 is positive for two microdissection clone markers, MD471 and MD412. MD412 gave a pattern of hybridisation identical to the left vector arm

probe, enabling positioning of this marker at a maximum distance of 120kb from the left hand end of the YAC to the nearest *Eag I* site. MD471 hybridised to the internal fragment where applicable, and to the right end fragment in other cases. This marker was therefore placed within a 70kb *Sfi I / BssH II* fragment (Fig. 3.10)

Figure 3.10 Long range restriction maps of ICI YACs 19EE2 and 8AB2. Shaded areas represent digest bands positive for MD122 (19EE2) or MD471 (8AB2). MD412 gave the same pattern of hybridisation as the left hand vector probe.



3.4.5 Discussion

Pulsed field maps of some other YACs in the contig were constructed in a similar manner. YACs D0485, A0846, D11155 and A04135 were mapped by John Brown, YACs 36IF5 and 16GB9 by Kathy Evans and YAC 14IA11 by John Maule (data not shown). Combining the results from these YACs with those for 8AB2 and 19EE2 enabled a long range restriction map to be produced across the whole YAC contig (data not shown). This map included all the microdissection clone markers depicted in Fig. 3.9. The exact length of the contig could not be determined due to the unknown degree of overlap between YACs, but could be estimated as a minimum of 3Mb, as defined by the minimal set of non-overlapping YACs D0485, D11155 and H0921.

The number of genes within the region could be estimated from the number of *Not I* sites within the YAC contig, since 89% of *Not I* sites occur within a CpG island (Lindsay and Bird, 1987). There were only three *Not I* sites within the entire contig, all of which occurred at the proximal end of YAC D0485. This suggests that the region contains few genes. However, only 12% of islands contain a *Not I* site (Lindsay and Bird, 1987), and only 40% of tissue specific genes are associated with a CpG island (Larsen *et al*, 1992), so this presumption must be viewed with caution. The presumed distribution of genes is consistent with the cytogenetic location of the contig, at the junction between a G (gene poor) an R (gene rich) band.

CHAPTER 4

A SEARCH FOR GENES IN THE BREAKPOINT REGION BY COINCIDENT SEQUENCE CLONING

Once it had been established that YAC D0485 spanned the translocation breakpoint it was important to begin to search for genes in this region, with a view to explaining the significance of the translocation event in relation to psychiatric illness. The choice of method for gene finding was coincident sequence cloning. Both a hybrid fishing CSC (HF-CSC) and an end ligation (EL-CSC) experiment were carried out. The hybrid fishing experiment and product analysis was carried out entirely by myself. The end ligation experiment was carried out by Tony Brookes, and then the product library was analysed by Kathy Evans, myself, Tony Brookes and Susan Anderson. This chapter is concerned with the hybrid fishing experiment only except where otherwise stated.

The input DNAs in the CSC experiment were YAC D0485 (genomic resource) and human foetal brain cDNA (cDNA resource) (obtained from Derek Blake, Institute of Molecular Medicine, Oxford). Both the genomic and the cDNA resources were catch linkered, so as to create a simplified resource, amplified via linker sequences appropriate for use in CSC.

4.1 Amplification of the Product cDNA

The two input DNA sources were integrated as in section 2.14.4, then after the washing steps the product cDNA was eluted. This product was amplified with primer 727 which is complementary to the linker sequences present on the ends of the cDNA molecules. The two control products, 'cDNA only' and 'no DNA' were also amplified with primer 727. The programme used was:
Denaturation: 93°C for 30 secs in the first cycle and 15 secs thereafter.
Annealing: 54°C for 30 secs.

Extension: 72°C for 1 min 30 secs (cycles 1-10), 2 mins 30 secs (11-20), 4 mins (21-29) and 8 mins (cycle 30).

The PCR products were run on a 2% agarose gel (Fig.4.1). The product cDNA track contained a smear of molecules larger than ~250bp, with one major product band at ~500bp. This band was excised from the gel for future use as a probe. In the 'cDNA only' track was a very faint smear. This probably represents the small amount of cDNA which adhered non-specifically to the beads and was not removed by the washing steps. The 'no DNA' track was completely blank. This implies that no DNA which could be amplified by primer 727 had been introduced to the experiment as a contaminant. The negative control for the PCR reaction (no template DNA included) was also blank.

4.2 Cloning of the Product cDNA

The linker used to amplify the cDNA incorporates an *EcoRI* site, which greatly facilitated the cloning step since cohesive end cloning could be employed. 20µl of the amplified product was digested with *EcoRI* and cloned into pBS vector which had been digested with *EcoRI* and phosphatased. 20ng of the digested cDNA was ligated into 20ng of pBS vector in a total of 20µl. The transformation efficiency was approximately 3000 white colonies per ng of product cDNA. The cells were plated at an optimum density, so that individual colonies could be recognised and accurately picked. All the colonies from several of these plates were lifted onto nitrocellulose filters, to provide high density (~1500 colonies) filters for estimation of the frequency of products. In addition 480 white colonies were picked at random and each grown in a well of a microtitre plate. In this way, a representative portion of the library was ordered onto five 12 x 8 grids, such that each colony could be referred to by its plate number and grid reference. The library was subsequently manipulated by a Hybaid colony picking robot, to make frozen stocks of the colonies and to replicate the grids onto nylon filters for hybridisation.

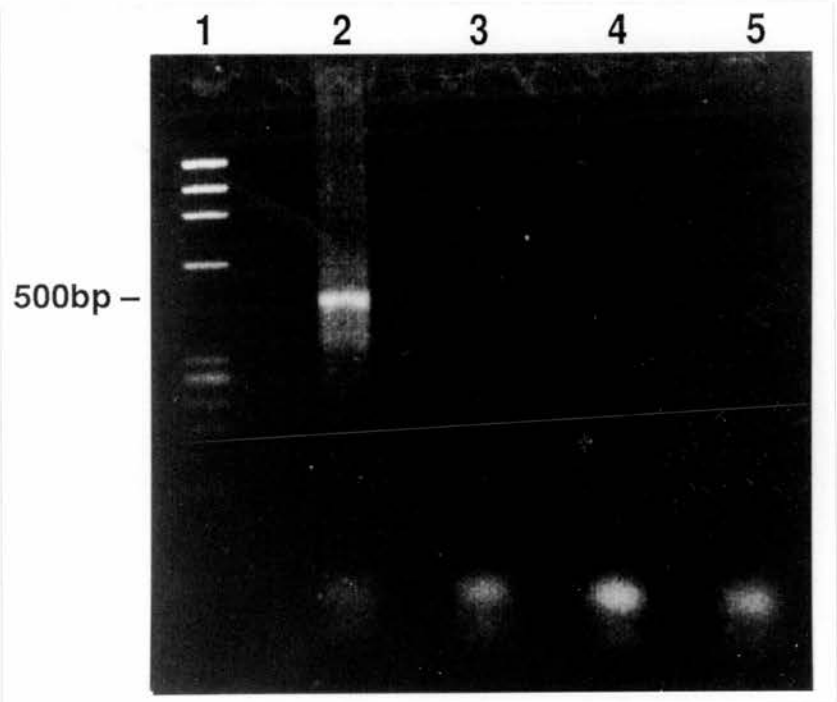


Figure 4.1 Amplified product cDNA from the HF-CSC experiment

The marker used is 250ng ϕx *Hae III* (track 1).

Tracks 2 to 5 show 6µl PCR product resolved on a 2% agarose gel. Track 2 contains the product cDNA, linkered and amplified with primer 727. A major product occurs at 500bp. Tracks 3 and 4 contain the CSC negative controls, 'cDNA only' and 'no DNA' respectively. Track 5 contains a 'no template' PCR negative control.

4.3 Analysis of the Product Library

4.3.1 Strategy for analysis of the library

A thorough analysis of the 480 clones comprising the gridded library was undertaken, with an exhaustive analysis of the first 192 clones. The first step was to identify by hybridisation clones containing high copy repeat elements or ribosomal DNA, so that these could be excluded from further analysis. Secondly, the major product was identified by using the PCR product band as a probe onto the gridded filters. Thirdly, representative positive clones were picked, and their insert amplified by PCR. Clones with a reasonably sized insert (more than 100bp) were sequenced by single pass automated cycle sequencing. The sequences at the end of the clones were checked to ensure they contained the expected sequence derived from the linker primer. Sequences were edited to remove vector and primer sequences and were then analysed by computer for sequence matches with already identified products and sequences in the Genbank, EMBL and dbEST databases. Probes were made from selected products by removing the vector sequences by *EcoRI* digestion and then excising the insert band from an LMP gel. All products identified were hybridised back onto the library to avoid reselection of previously analysed products. The identity of representative clones positive for a probe by hybridisation was verified by sequencing. Clones whose identity had not been established by hybridisation were picked at random, and those with an insert of reasonable size were sequenced.

4.3.2 Assessment of the frequency of high copy repeats and ribosomal DNA sequences in the library

The high frequency of repetitive elements in both genomic DNA (Alu elements comprise approximately 5% of the human genome; Shen *et al*, 1991) and cDNA (3-5% of cDNAs contain Alu elements; Crampton *et al*, 1981) means that genuine recovery of these repeats by coincidence can be problematic. Clones which contain an Alu repeat are difficult to use as hybridisation probes, since the signal resulting from the repeat will obscure

that resulting from the non-Alu portion of the probe. Before integration, the denatured genomic DNA and cDNA had been allowed to anneal to an excess of Human Cot1 DNA, which is comprised almost entirely of high copy repeat sequences, primarily Alu elements. It was hoped that this pre-annealing step would 'quench' most of the Alu elements and prevent them from occurring in the product DNA.

In order to estimate the frequency of clones containing Alu elements in the product library, seven high density filters were screened with human Cot1 DNA. There were 44 positive clones, which gives a frequency of 0.4% clones containing Alu elements. This frequency was sufficiently low that the presence of high copy repeat elements would not hamper the analysis of the product library. The five gridded filters were also screened with Cot1 DNA. Six colonies out of the 480 gave strong positive signals. Two of these were sequenced and they were found to be identical to each other, and to have high homology with Alu-related repetitive elements present in known human genes.

The presence of a high frequency of ribosomal DNA clones in the product library was also a potential problem since there is a very high frequency of ribosomal sequences in yeast DNA, which are sufficiently conserved with their human counterparts that they may be recovered by coincidence. Ribosomal DNA from two clones had been used to pre-anneal to the input DNAs in order to block their involvement in IRD formation. The same clones were labelled and used as a probe onto the product library, to assess the efficiency of blocking. There were 52 positive clones on 6 high density plates, which implies a frequency of 0.6%. This was deemed adequately low. Only 2 clones from the grids gave a strong signal with the ribosomal DNA probes. No further analysis was performed on these clones.

4.3.3 Identification of the major product

The one band resulting from the initial PCR reaction was likely to comprise a product at high frequency in the library, and therefore was used as a probe onto the gridded filters.

This probe hybridised to approximately 10% of the colonies (5% if only the strong signals were included). Twenty clones which hybridised strongly were picked, and the insert amplified by PCR. Five of these, which all had different insert sizes, were chosen for sequencing. These were clones 1-6h, 2-2b, 2-10b, 3-6c and 1-10d.

Out of these five clones, four (excepting 2-10b) gave highly significant database homology matches with parts of an α-tubulin gene. 2-2b (see figure 4.2) and 1-10d matched 5' portions of the gene, 1-6h matched the 3' end (see figure 4.3), and 3-6c matched a more central section. The α-tubulin clones 1-6h, 2-2b and 3-6c were used as probes onto the gridded library. Almost all the clones which had been positive for the major PCR band probe were positive for either 1-6h or 2-2b, confirming that the major product of the HF-CSC was α-tubulin. There was no overlap in hybridisation pattern between these two probes. There were no clones positive for 3-6c, except itself.

The hybridisation of the PCR product probe to clone 2-10b was strong and unlikely to have been non-specific. The PCR band probe was therefore likely to have contained 2-10b sequences in addition to α-tubulin. The sequence of this clone gave no significant matches in the databases. Three other products were identified from the gridded library whose sequence matched 2-10b. Further analysis on these clones will be described in section 4.6.

Figure 4.2 Sequence match between the CSC clone 2-2b and the keratinocyte alpha tubulin gene ('ker').

```
2-2b TCGACTCTTAGCTTGTCGGGGACGGTAACCGGGACCCGGTGTCTGCTCCT  50
                  |||||||||||||||||||||||||   |||||||||
ker               TGTCGGGGACGGTAACCGGGACCCGTGCTCTGCTCCT  50

2-2b GTCGCCTTCGCCTCCTG-ATCCCTAGCCACTATGCGTGAGTGCATCTCCA 100
     |||||||||||||||| |||||||||| ||||||||||||||||||||
ker  GTCGCCTTCGCCTCCTGAATCCCTAGCCA-TATGCGTGAGTGCATCTCCA 100

2-2b TCCACGTTGGCCAGGCTGGTGTCCAGATTGGCAATGCCTGCTGGGAGCTC 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
ker  TCCACGTTGGCCAGGCTGGTGTCCAGATTGGCAATGCCTGCTGGGAGCTC 150

2-2b TACTGCCTGGAACACGGCATCCAGCCCGATGGCCAGATGCCAAGTGACAA 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
ker  TACTGCCTGGAACACGGCATCCAGCCCGATGGCCAGATGCCAAGTGACAA 200

2-2b GACCATTGGGGGAGGAGATGACTCCTTCAACACCTCCTTCAGTGAGACGG 250
     |||||||||||||||||||||||||||||||| ||||||||||||||
ker  GACCATTGGGGGAGGAGATGACTCCTTCAACACCTTCTTCAGTGAGACGG 250

2-2b GCGCTGGCAAGCACGTGCCCCGGGCTGTGTTTGGTAGACTTGGTACCCAC 300
     |||||||||||||||||||||||||||||| |||||||||| ||||||
ker  GCGCTGGCAAGCACGTGCCCCGGGCTGTGTTT-GTAGACTTGGAACCCAC 300

2-2b AATCATTGATGAAGTTCGCACTTGGCAACTACCGCCAGCTCTTCCACCCT 350
     | ||||||||||||||||||| ||||| ||||||||||||||||||||||
ker  AGTCATTGATGAAGTTCGCAC-TGGCACCTACCGCCAGCTCTTCCACCCT 350

2-2b GAGCAGCTCATCACAGGCAAGGAAGATGCTGCCAATAACTATGCCCGAGG 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
ker  GAGCAGCTCATCACAGGCAAGGAAGATGCTGCCAATAACTATGCCCGAGG 400

2-2b GCACTACACCATTGGCAAGGAGATC  425
     |||||||||||||||||||||||||
ker  GCACTACACCATTGGCAAGGAGATC  425
```

Figure 4.3 Sequence match between the CSC clone 1-6h and the foetal brain alpha tubulin gene ('fbrn').

```
1-6h TAAATACATGGCTTGCTGCCTGTTGTACCGTGGTGACGTGGTTCCCAAAG    50
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
fbrn TAAATACATGGCTTGCTGCCTGTTGTACCGTGGTGACGTGGTTCCCAAAG  3491

1-6h ATGTCAATGCTGCCATTGCCACCATCAAGACCAAGCGTACCATCCAGTTT   100
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
fbrn ATGTCAATGCTGCCATTGCCACCATCAAGACCAAGCGTACCATCCAGTTT  3541

1-6h GTGGATTGGTGCCCCACTGGCTTCAAGGTTGGCATCAACTACCAGCCTCC   150
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
fbrn GTGGATTGGTGCCCCACTGGCTTCAAGGTTGGCATCAACTACCAGCCTCC  3591

1-6h CACTGTGGTGCCTTGATGGAAGACCTGGCCAAGGTACAAGAGAgCTGTGT   200
     |||||||||||| || ||| ||||||||||||||||| ||||||||||||
fbrn CACTGTGGTGCC-TGGTGG-AGACCTGGCCAAGGTAC-AGAGAGCTGTGT  3641

1-6h GCATNCTGAgCAAACACCACAACCATTGCTGAGGCCTGGGCTCGCCTGGA   250
     |||| |||||| |||||||| ||||||||||||||||||||||||||||
fbrn GCATGCTGAGC-AACACCACAGCCATTGCTGAGGCCTGGGCTCGCCTGGA  3691

1-6h CCACAAGTTTGACCTGATGTATGCCAAACGTGCCTTTGTTCACTGGTACG   300
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
fbrn CCACAAGTTTGACCTGATGTATGCCAAACGTGCCTTTGTTCACTGGTACG  3741

1-6h TTGGGGAGGGGATGGAGGAAGGTGAGTTTTCAGAGGCCCGTGAGGACATG   350
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
fbrn TTGGGGAGGGGATGGAGGAAGGTGAGTTTTCAGAGGCCCGTGAGGACATG  3791

1-6h GCTGCCCTTGAGAAGGATTATGAGGAGGTTGGTGTGGATTCTGTTGAAGG   400
     |||||||||||||||||||||||||||||||||| ||||||||||||
fbrn GCTGCCCTTGAGAAGGATTATGAGGAGGTTGGTGTGCATTCTGTTGAAGG  3841

1-6h AGAGGGTGAGGAAGAAGGAGAGGAATACTAAAGTTAAAACGTCACAAAGG   450
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
fbrn AGAGGGTGAGGAAGAAGGAGAGGAATACTAAAGTTAAAACGTCACAAAGG  3891

1-6h TGCTGCTTTTACAGGGAAGCTTATTCTGTTTTAAACATTGAAAA-GTTGT   500
     |||||||||||||||||||||||||||||||||||||||||| |||||
fbrn TGCTGCTTTTACAGGGAAGCTTATTCTGTTTTAAACATTGAAAATGTTGT  3941

1-6h GGTCTGATC   509
     |||||||||
fbrn GGTCTGATC  3950
```

In addition to the five clones positive for the PCR band, and the two Alu repeat-containing clones, a further 68 products from the gridded library were sequenced. Of these, 50 were identified as parts of an α-tubulin gene. Three clones gave significant database matches with other known genes, the non-erythroid spectrin gene, the neurodevelopmental gene neuronatin, and a ribosomal protein gene. One was a short interspersed repeat element (SINE). Three gave sequence matches with 2-10b (see above) and a further 11 clones had no matches in the database.

4.3.5 A complete analysis of a portion of the library

A virtually complete analysis of 40% of the gridded library was undertaken. The insert of 188 of the 192 clones was amplified by PCR and the products run out on 1.5% agarose gels. It was found that of these, 95 of the clones (50%) possessed either no insert or one too small for useful analysis (less than ~100bp). Clones consisting of purely vector sequence could have resulted from incomplete digestion of the vector, or incomplete phosphatase treatment (which stops the vector from recircularising during the ligation step). Small insert fragments will be cloned preferentially to larger ones, so this could explain the high proportion of small clone inserts. Larger clone insert sizes could have been obtained by using a gel purification step to size select the PCR amplified material to be cloned.

Of the remaining 93 clones, the average insert size was approximately 500bp. One colony gave two bands upon PCR with the vector primers flanking the insert. This could reflect a colony composed of two different clones, a co-cloning of two different plasmids into the same bacterial cell or a complex chimaeric plasmid formed between two vector and insert sequences. For twelve clones, the miniprep or sequencing reaction failed upon the first attempt, and was not repeated. Sixty-nine clones (36%) were identified as being parts of an α-tubulin gene, 35 of these by sequence analysis and 34 by hybridisation with other α-tubulin clone probes only. Two

(1%) were positive with the human Cot1 DNA probe and therefore contained high copy repeats. Two products gave database matches to known genes other than α-tubulin and five products had no sequence matches in the database (4% non-tubulin gene products). Of the latter seven, three were later found to be artefactual products, resulting from amplification and cloning of non-coincident cDNA. This implies that less than 3% of the cloned CSC products were artefactual (3 of the 93 clones with a reasonably sized insert).

Table 4.1 summarises the main features of the HF-CSC library (as estimated by analysis of the first 40% of the 480 colonies in the gridded library)

Table 4.1

Clone type	Frequency (%)
No insert / insert < 100bp	50
high copy repeat containing	1
ribosomal DNA	0.4
α-tubulin gene fragments	36
other known genes	1
no database matches	3
unidentified (sequencing failure)	9

4.4 Clones Containing Fragments of an Alpha Tubulin Gene

4.4.1 Tubulin Genes

There are currently three classes of known tubulin genes, alpha, beta and gamma. Gamma tubulin is associated with the centrosome (Zheng *et al*, 1991). Alpha and beta tubulin together comprise the major soluble protein component of microtubules, and are members of small multigene families in higher eukaryotes. Southern blot analysis reveals that there are approximately fifteen copies of both the alpha and beta tubulin sequence in the human genome, though not all of these copies are necessarily

expressed. Extensive analysis on the beta tubulin gene family has revealed that most of these bands correspond to non-functional pseudogenes (Cleveland and Sullivan, 1985), and the same is likely to be true for alpha tubulin.

For alpha tubulin, three expressed sub-types have been identified. The first, foetal brain alpha tubulin, demonstrates a pattern of expression restricted to differentiated cells of neurological origin (Cowan *et al*, 1983). The second has a more ubiquitous expression pattern, and is termed keratinocyte alpha tubulin, reflecting the cell type from which it was isolated (Cowan *et al*, 1983). The third, testis specific alpha tubulin, is only expressed in the testis of monkeys and humans, and has been mapped to chromosome 2q (Gerhard *et al*, 1985).

The structure of both foetal brain and keratinocyte α-tubulin comprises 4 exons and 3 introns, the first exon consisting of just the initiator methionine codon. All three contain a poly(GT) repeat in the first intron (Todd and Naylor, 1991, Hall and Cowan, 1985). There is 97% sequence homology between these two alpha tubulin genes, but the 5' and 3' untranslated regions (UTRs) are different (Cowan *et al*, 1983). There is however considerable inter-species homology (65-80%) in the non-coding regions, which suggests selective conservation throughout evolution, possibly due to an as yet unknown functional requirement.

The high sequence similarity between the coding regions of the foetal brain and keratinocyte α-tubulin genes means that hybridisation with these portions of the gene cannot distinguish between the two. Probes must be designed from the UTR sequences. Hybridisation of a UTR-derived probe from both subtypes to total human DNA detects two bands, which are specific to the subtype. This suggests that only four out of the fifteen α-tubulin gene copies contain sequence similarity with two of the three known subtypes.

The chromosomal loci for the foetal brain or keratinocyte α-tubulin genes are unknown. Loci for α-tubulin genes or pseudogenes have only been detected on chromosomes 2q (testis specific gene, TUBA1, Gerhard *et al*, 1985), 13q11 (tubulin alpha-like 2, Bonaldo, 1992) and 12 (tubulin alpha-like 1, Gatti *et al*, 1987). In addition some bands produced by digestion of the chromosome 12 family member were identified as present on chromosome 11 (Gatti *et al*, 1987).

4.4.2 Identification by sequencing and hybridisation

A total of 128 clones containing α-tubulin gene fragment inserts were identified from the whole gridded library of 480 clones (26%). The complete analysis of 40% of the library suggested that 36% of the clones were α-tubulin, which implies that more α-tubulin clones would have been identified had a complete analysis of the whole gridded library been undertaken.

Of these 128 clones, 54 were proven to be fragments of the α-tubulin gene by sequence analysis. The other 74 were positive by hybridisation only. These 74 clones gave strong positive signals with clone 2-2b, clone 1-6h or tubulin clone 'random4' from the EL-CSC experiment and/or the PCR band probe and/or a composite α-tubulin probe made up of cDNA fragments spanning as much as possible of the whole gene (comprising the inserts of clones 2-2b, 1-6h, random4 and 1-f5). All these probes, except the PCR band, had been proven by sequencing to contain α-tubulin fragments alone, hence it is reasonable to assume that clones unambiguously hybridising to them do indeed contain tubulin sequences. Only seven clones were positive by hybridisation to the PCR band only. There must be some doubt as to whether these clones definitely contain α-tubulin sequence, however none were positive for any other probes.

4.4.3 Comparison of CSC α-tubulin clones with the known α-tubulin gene sequences

Figure 4.4 shows the position of the α-tubulin clones isolated from the CSC experiment in relation to the known expressed α-tubulin genes from foetal brain and keratinocyte. It can be seen that a virtually complete coverage of the coding sequence was achieved. The only portion missing was 130bp from exon 4. This could be because it was not recovered in the CSC product DNA, or because it was refractory to cloning. An 110bp fragment from exon 4 was present most abundantly - 29 of the 54 α-tubulin clones sequenced were derived from this region.

The foetal brain and keratinocyte α-tubulin cDNA sequences are sufficiently similar throughout most of the translated region that it is impossible to tell which of the individual cDNA clones were derived from each. Single pass ABI cycle sequencing is only 98-99% accurate, and so small variations from the expected sequence may be attributed to sequencing errors. Only in a small region of exon four is there sufficient divergence between the two subtypes (a block of nineteen nucleotides in which there are ten mismatched bases) that assignment to a particular α-tubulin could be accomplished. It was found that cDNA clones from both types had been isolated. Slightly more clones in this region were derived from foetal brain cDNA (six clones) than from keratinocyte cDNA (four clones). This is unsurprising considering that the cDNA source for the CSC experiment was foetal brain.

Clones in which the sequence extended into the 5' or 3' untranslated (UTR) regions were easily assigned to be derived from a particular expressed α-tubulin, since the UTR sequences of the known expressed alpha tubulin subtypes show no homology among themselves. Of the four clones which extended into the 3' UTR, three were derived from foetal brain cDNA and two from a keratinocyte-like cDNA. Four clones extended into the 5' UTR (one of these from the EL-CSC experiment). Of these, two were derived from a

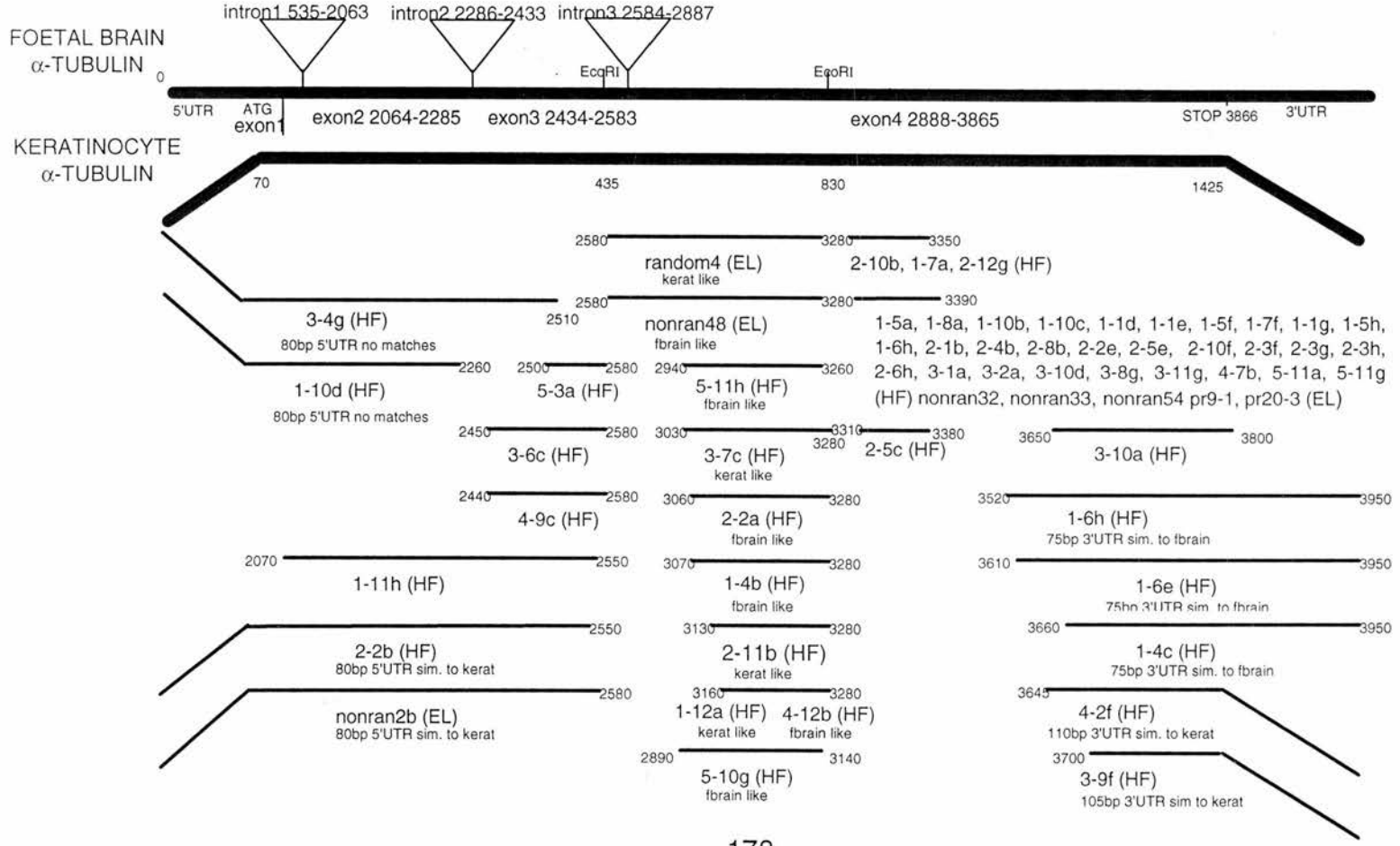
keratinocyte-like cDNA and the other two did not match any known tubulin. The suffix 'like' is added above since the UTR's of the keratinocyte cDNA clones did not match perfectly with the published keratinocyte α-tubulin sequence. The rate of mismatch was slightly higher than could be attributed to normal sequencing errors - for instance for the clone 2-2b there were 5 mismatches with the keratinocyte α-tubulin within the 70bp of the UTR, but then only a further 6 mismatches in the remaining 270bp of the clone. Similarly for the EL clone nonran2b there were 10 mismatched bases in the 70bp of the UTR, and only 9 in the remaining 340bp of the clone. This implies that there may be other expressed α-tubulin sequence similar to keratinocyte α-tubulin. The UTR sequences of 2-2b and nonran2b are however not identical, which implies either the existence of more than one additional keratinocyte-like expressed α-tubulin gene or a chance excess of sequencing errors in the UTRs.

Two of the clones with sequence extending into the 5' UTRs (1-10d and 3-4g) were identical to each other, but did not match either the foetal brain or the keratinocyte α-tubulin genes in the UTR region. This implies that there is definitely a third type of expressed tubulin gene. No further analysis of this gene was performed.

4.4.4 An α-tubulin gene on YAC D0485

Three of the CSC α-tubulin clones were used as hybridisation probes back to the starting genomic resource, that is YAC D0485. The positive hybridisation signal confirmed that there was genuinely at least one α-tubulin sequence located on the YAC (see Fig. 5.1). Chapter 5 is concerned with the finer mapping and further analysis of this sequence.

Fig.4.4 Alignment of CSC α-tubulin clones along the known foetal brain (upper thick black line) and keratinocyte (lower thick black line) expressed α-tubulin genes. Positions of the introns in foetal brain α-tubulin are marked. The 5' and 3' UTR regions are shown at different angles, since there is no similarity in sequence between the two genes in these regions. The angle at which the UTR regions from the CSC clones represents the α-tubulin subtype from which the clones were derived. Two clones were derived from a novel type of expressed α-tubulin (3-4g and 1-10d). Alpha tubulin clones from the hybrid fishing (HF) and the end ligation (EL) CSC experiments are shown for completeness. Numbers at each end of the clones represent their position on foetal brain α-tubulin.



4.5 Clones derived from other known genes

Three clones identified from the gridded library matched parts of known genes other than α-tubulin.

4.5.1 A ribosomal protein

The sequence of one clone (1-7e) was 97% identical to part of the human ribosomal protein L7a large subunit mRNA. This ribosomal protein gene had previously been assigned to chromosome 9q34 (Yon *et al*, 1989). Clone 1-7e may have been recovered by coincidence, indicating a locus very similar to that of the ribosomal protein on chromosome 11q, or alternatively it could have been recovered as an artefact. Hybridisation of the insert of this clone onto YAC D0485 (restriction digest fragments resolved on a pulsed field gel) gave no signal at normal stringency and thus it was interpreted that this clone represented an artefactual product.

5.5.2 Neuronatin

Clone 4-8c gave significant matches to both rodent genes and human EST's. The same portion of the clone produced a 73% identity to the *Rattus Norvegicus* neuronatin gene, alpha and beta mRNAs (see figure 4.5a). The gene neuronatin was identified by differential display to define genes expressed in postnatal brain development (Joseph *et al*, 1994). A full length cDNA was isolated, which coded for a novel protein of 81 amino acids, which was expressed highly in both rat neonatal and human foetal brain. The gene was expressed during late foetal and early postnatal brain development and was postulated to be involved in terminal brain differentiation (Joseph *et al*, 1994).

The same part of the CSC clone also produced a 71% identity to rat brain calbindin protein 28K. Upon closer inspection of the calbindin sequence in the database it was discovered that the neuronatin gene sequence is joined

on after the calbindin sequence itself, and is termed an 'unrelated cDNA fragment'. The match to calbindin was therefore an artefact of the database.

A search for homology with human EST's produced two extremely close matches (201 bases identical out of 203 for human EST 135164 and 230 bases identical out of 231 for human EST 13968) (see figure 4.5b). With near-perfect matches such as this it can be assumed that these two EST's and the clone 4-8c are derived from the same cDNA species.

The fact that the gene homologous to clone 4-8c, neuronatin, is involved in brain development made it a very attractive candidate for involvement in schizophrenia. However, no signal could be seen upon hybridisation of the insert of this clone onto YAC D0485, even under reduced stringency conditions (hybridisations and washing at 55°C). It was therefore concluded that this clone was an artefact in the CSC library.

Figure 4.5a Sequence match between CSC clone 4-c8 and the rat brain development gene neuronatin ('neuro').

```
4-c8   GGAAATCCAAAACACCGCACCAGCCAGCAGGAATGGACATTCTGACATCG    50
           |     |||||| ||||| |||||| ||||||||||||||||||||
neuro  ATGGAGAGCTAACACCCCACCAACCAGCA-GAATGGACATTCTGACATCA  683

4-c8   CCAGCCGACGCCCTGAATCTTTGGTGCAGCA--------CCCAACCGCGT  100
       ||||| ||  |||||||||| | ||||||| |          |||| ||||
neuro  CCAGCTGAAACCCTGAATC-TCGGTGCAGAAGAGAAAGTGTCAACTGCGT  733

4-c8   GCCTGTGTGGCGGGACTGGAGGGCACAGTTGAGGAAGGAGGGTGGTTAAG  150
       ||     ||| |||| |||||||    | ||  | |||| |  |||||||
neuro  GCAGCACTGG-GGGAGTGGAGGGTGTGGGTGGTGGAGGAAGAGGGTTAAG  783

4-c8   AAATACAGTGGGGCCCTCTCGCTGTCCCTTGCCCAGGGCACTTGCATTCC  200
       |||   |||||||||||||| |||||||||||| | |||||    |||||
neuro  AAAACTAGTGGGGCCCTCTTGCTGTCCCTTGCCTATGGCACGCATATTCC  833

4-c8   AGCCTCGCTGCATTTGCTCTCTCGATTCCCCTTTCCTCCTCACTGCCTCC  250
        |||          |||||| |||  | ||||| ||| |||  ||   |||
neuro  TGCC---------TTGCTCCCTCACTCCCCCTCTCC-CCTGCCT---TCC  883

4-c8   CAAGCCCACCCTACTCCAAAATAATGTGTCACTTGATTTGGAACTATTCA  300
        ||||||  ||  | |||||  |||||||||||||||| ||| |||||||
neuro  AAAGCCCCACC--CCCCAAA--AATGTGTCACTTGATTCGGACCTATTCA  933

4-c8   AGCAGTAAAAGTAAATGAATC  321
       | ||||||  | |
neuro  ACCAGTAATTGGATCCCACCT  954
```

Figure 4.5b Sequence match between the CSC clone 4-c8 and the human EST sequence 13968 ('est').

```
4-c8 GGAAATCCAAAACACCGCACCAGCCAGCAGGAATGGACATTCTGACATCG   50
                                                ||||||||||||||||
est                                             GGACATTCTGACATCG   50

4-c8 CCAGCCGACGCCCTGAATCTTTGGTGCAGCACCCAACCGCGTGCCTGTGT  100
     |||||||||||||||||| |||||||||| ||||||||||||||||||
est  CCAGCCGACGCCCTGAATC-TTGGTGCAGCA-CCAACCGCGTGCCTGTGT  100

4-c8 GGCGGGACTGGAGGGCACAGTTGAGGAAGGAGGGTGGTTAAGAAATACAG  150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
est  GGCGGGACTGGAGGGCACAGTTGAGGAAGGAGGGTGGTTAAGAAATACAG  150

4-c8 TGGGGCCCTCTCGCTGTCCCTTGCCCAGGGCACTTGCATTCCAGCCTCGC  200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
est  TGGGGCCCTCTCGCTGTCCCTTGCCCAGGGCACTTGCATTCCAGCCTCGC  200

4-c8 TGCATTTGCTCTCTCGATTCCCCTTTCCTCCTCACTGCCTCCCAAGCCCA  250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
est  TGCATTTGCTCTCTCGATTCCCCTTTCCTCCTCACTGCCTCCCAAGCCCA  250

4-c8 CCCTACTCCAAAATAATGTGTCACTTGATTTGGAACTATTCAAGCAGTAA  300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
est  CCCTACTCCAAAATAATGTGTCACTTGATTTGGAACTATTCAAGCAGTAA  300

4-c8 AAGTAAATGAATC  313
     |||||||||||||
est  AAGTAAATGAATC  313
```

4.5.3 Non-erythroid spectrin

Clone 2-12f produced a 96% identity (208 out of 210 bases) with the human non-erythroid alpha spectrin gene. This gene, also termed alpha fodrin, has been mapped to 9q33-q34 (Upender *et al*, 1994). The very high level of identity between the clone and the α-fodrin gene implies that they were derived from the same cDNA (not another transcribed from a locus on 11q). The existence of an α-fodrin pseudogene within YAC D0485 could have enabled the selection of the α-fodrin cDNA; however there was no signal upon hybridisation of clone 2-12f to YAC D0485 under normal high stringency and so it is likely that this clone represents an artefactual product.

4.5.4 Discussion

Since these three clones did not hybridise back to the starting genomic resource, they probably represent artefactual products in the library. The stringency of washes used in the CSC experiment (0.1x SSC at 68°C) was the same as those used in the hybridisation to D0485, so it is unlikely that these products could have been recovered by genuine coincidence, but with imperfect base-pairing such that they did not hybridise to D0485.

It is interesting however, that rôles can be foreseen for both alpha-fodrin and neuronatin in psychiatric illness; the latter because it is involved in brain development and the former because alpha-fodrin is part of a super family of structural proteins (Davison *et al*, 1989), of which itself and other members may be of significance to the (1;11) translocation and/or to schizophrenia. Alpha-fodrin is capable of binding actin, and in myelinated neurons may be responsible for the distribution of sodium channels at the nodes of Ranvier (Moon and McMahon, 1990). It is probably also important in the formation and maintenance of polarised membrane domains (Nelson and Veshnock, 1987). Other proteins in the spectrin superfamily include alpha-actinin and dystrophin. There are loci for alpha-actinin at 1q42-q43 (ACTN2) and 11q13-q14 (ACTN3) (Beggs *et al*, 1992), which are cytogenetically close (plus or

minus 10Mb) to the positions of the breakpoints of the translocation. Also dystrophin has been previously implicated in schizophrenia (see section 1.10.3). The possibility that structural proteins are involved in the aetiology of psychiatric illness will be discussed more fully in Chapters 5 and 7. Although no rôle for a ribosomal protein as a candidate gene in psychiatric illness can easily be foreseen, it is located within the same cytogenetic region of the genome as alpha-fodrin, that is 9q33-34.

4.6 Clones Derived From Novel cDNAs

Fifteen clones from the HF-CSC library gave no significant matches with sequences in the databases, and therefore were derived from novel cDNAs. These clones were 1-3c, 1-10d, 1-11d, 2-10b, 2-5g, 3-3a, 3-4e, 3-5g, 4-3b, 4-7a, 4-10a, 4-7d, 4-1f, 4-7g and 5-3c. No further analysis was performed on 1-10d and 4-3b since they both had insert sizes of less than 100bp.

Computer assisted comparison of the sequences of the remaining clones revealed that some could be arranged into families. The clones usually represented different *Sau3AI* fragments of the same cDNA species, although occasionally partial digestion led to overlapping restriction fragments. During amplification of the CSC product cDNA, the primer often annealed non-specifically, and in these cases the ends of the clones were not *Sau3AI* sites. This was in fact advantageous for analysis of the products, since the clones were thus often overlapping, and contigs of cDNA clones could be assembled.

Some of the clones identified from the EL-CSC product library matched the novel cDNA clones from the HF library. Some were identified by random sequencing (by Kathy Evans) and others were identified by hybridisation with HF product probes. The identity of some of those in the latter group was confirmed by sequencing (by Susan Anderson). In addition, some EL clones were used as hybridisation probes to identify additional family members in

the HF library. Incorporating the results from analysis of the EL library, the clones derived from novel cDNAs could be grouped into five families (Figure 4.6 and Figure 4.7).

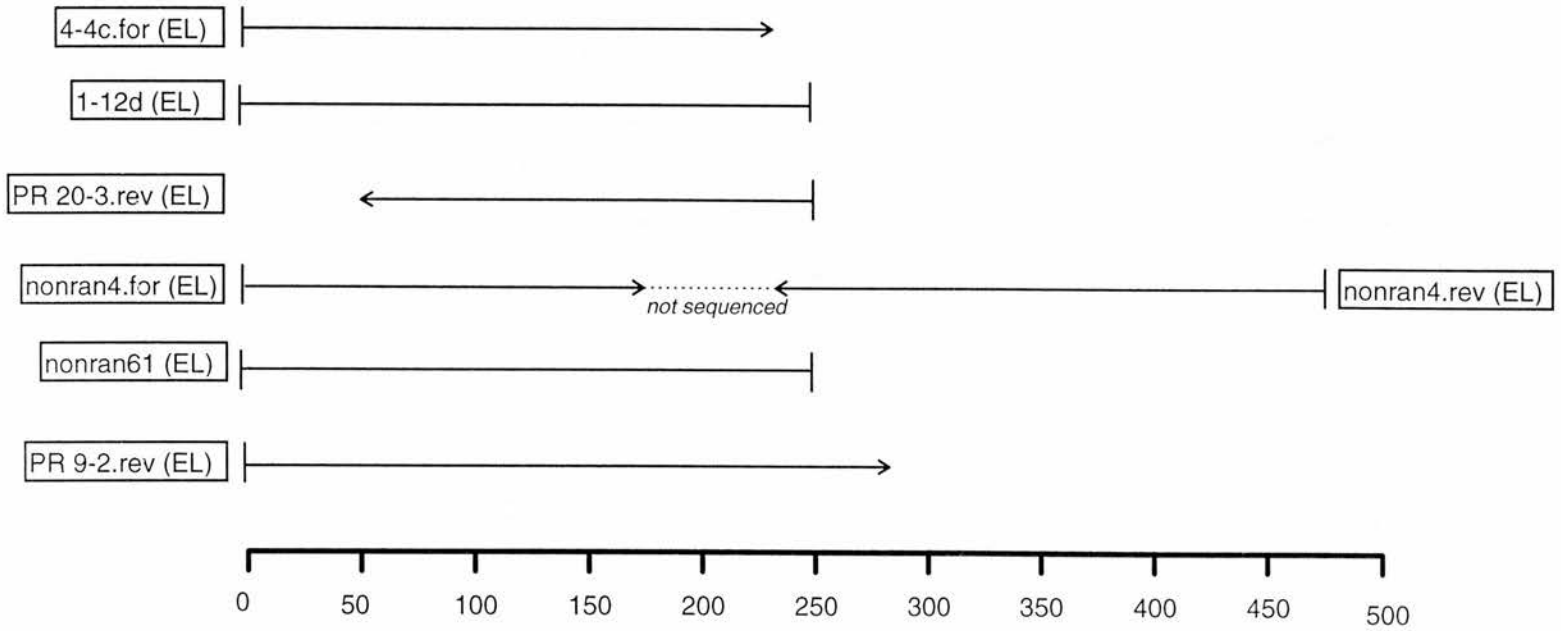
Family 1 was composed of sequences derived entirely from the end ligation experiment. The other families comprised HF and EL clones. Ten other probable Family 1 clones, eleven probable Family 2 clones and three probable Family 4 clones were identified from the EL library by hybridisation but were not sequenced. They have therefore not been included in Figures 4.6 and 4.7. The consensus sequences of the five families were then used to screen the databases for matches with known genes. No highly significant database matches were found.

Some of the CSC clones derived from novel cDNAs could not be grouped into a family. These were clones 2-5g, 4-1f and 3-3a, all derived from the HF experiment. Clone 2-5g was found to be an artefact of the library since it did not hybridise back to YAC D0485 under normal conditions of high stringency. Clones 4-1f and 3-3a did however hybridise to the YAC (see Fig. 6.2), these were termed Products 6 and 7 respectively. A small portion of clone 4-1f was 90% identical to part of the Alu repeat consensus sequence.

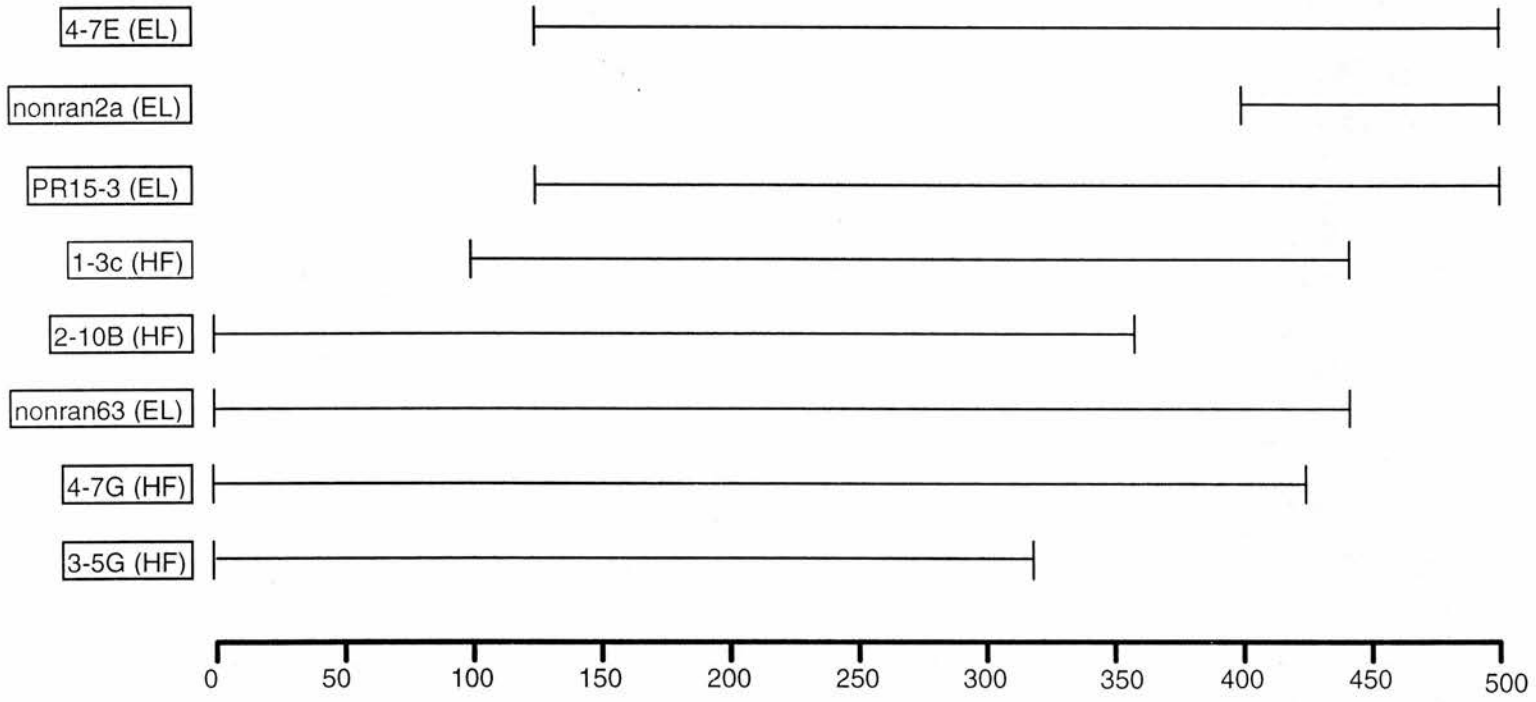
Although the coincident clones derived from novel cDNAs had been grouped into five families and two other single products, this implied nothing regarding the number of genes from which they were derived. All the products could be different fragments of the same gene or there could be several genes. Also some of the products may be low copy repeat sequences, artefactual though recovered by genuine coincidence. Further studies to prove the genic nature of the products and to estimate the number of genes from which they were derived are described in Chapter 6.

Fig.4.6 A)-E) Schematic diagram depicting alignment of CSC clones derived from novel cDNAs into five families based on sequence match. Hybrid fishing CSC (HF) and end ligation CSC (EL) clones are shown. The scale represents the length of the clones in base pairs. Clones which have not been sequenced across their whole length are shown with an arrow extending towards the non-sequenced portion. In these cases the clone has been sequenced with either primer 292 ('for') or 291 ('rev').

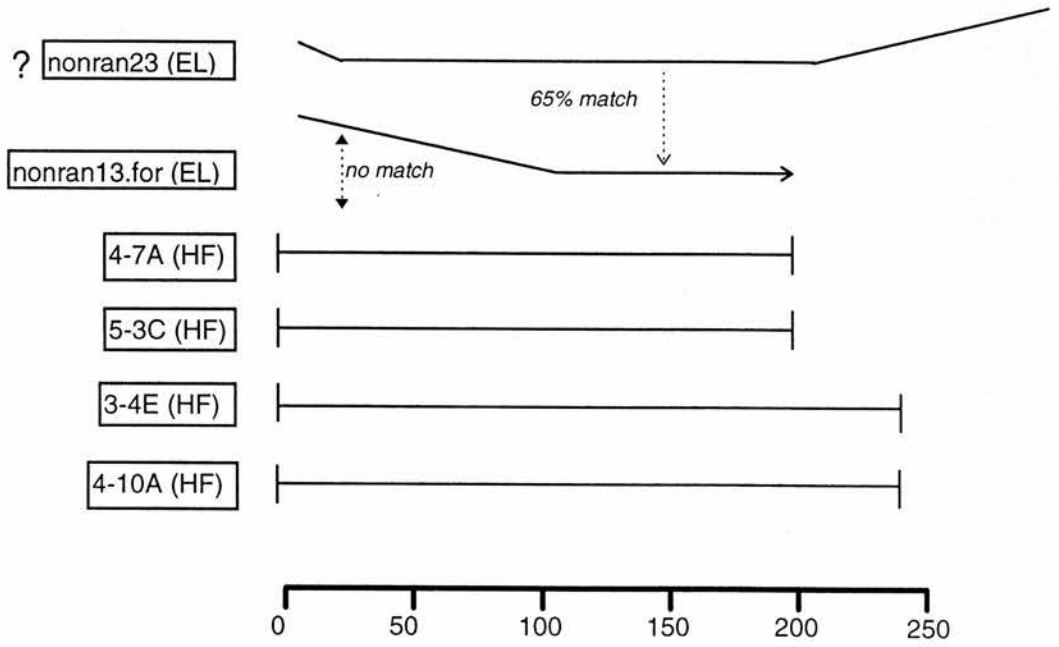
A) Family 1 is composed entirely of clones from the EL-CSC experiment. Nonran4 and nonran61 were isolated by random sequencing of clones. 4-4c and 1-12d were isolated by hybridisation with nonran4 and nonran61 probes. PR 20-3 was isolated with an α-tubulin clone probe (1-f5). PR 9-2 was isolated with EL PCR band 4.



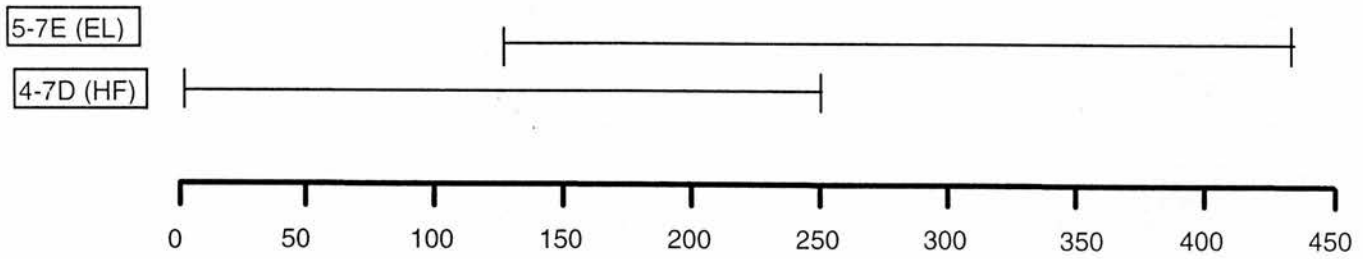
B) Family 2. 2-10b was isolated by hybridisation with the HF PCR band. 4-7g was isolated by hybridisation with 2-10b. Nonran2a and nonran63 were isolated by random sequencing. 4-7e, 1-3c and 3-5g were isolated by hybridisation with nonran2a and nonran63 probes. PR15-3 was isolated by hybridisation with the EL PCR band 6.



C) Family 3. 5-3c, nonran13 and nonran23 were isolated by random sequencing. 3-4e, 4-7a and 4-10a were isolated by hybridisation with a 5-3c probe. Nonran23 has only a 65% sequence match with the other products and is thus tentatively included in this family.



D) Family 4. 4-7d was isolated by random sequencing. 5-7e was isolated by hybridisation with 4-7d probe.



E) Family 5. 1-11d was isolated by random sequencing. 2-5c was isolated by hybridisation with 1-11d probe.

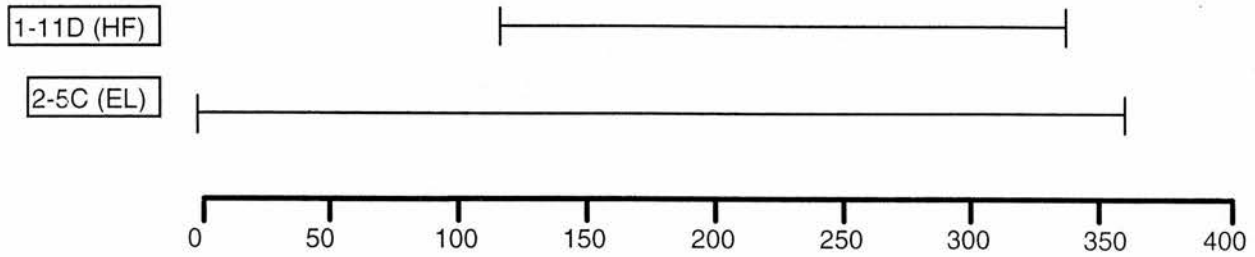


Figure 4.7 A) to G) Sequences of clones comprising Families 1 to 7 respectively. The consensus sequence ('CON') is also shown where the Family is comprised of more than one clone (i.e. Families 1 to 5 inclusive). Spaces inserted into the sequence in order to preserve alignment are shown by '.'

A) Family 1

'nr' = nonran, '9-2' = pr9-2, '20-3' = pr20-3. Nonran4 is incompletely sequenced; the non-sequenced portion is shown by '--'

```
nr4     AGGAGAAGCTCGTGGCTGCTGAGTTCTCCTGGCCACCATGAACTYCAGGA   50
nr61    AGGAGAAGCTCGTGGCTGCTGAGTTCTCCTGGCCACCATGAACTTCAGGA   50
9-2     AGGAGAAGCTCGTGGCTGCTGAGTTCTCCTGGCCACCATGAACTTCAGGA   50
4-4c    AGGAGAAGCTCGTGGCTGMTGAGTTCTCCTGGCCACCATGAACTTCAGGA   50
1-12d   AGGAGAAGCTCGTGGCTGCTGAGTTCTCCAGGCCACCATGAACTTCAGGA   50
CON     AGGAGAAGCTCGTGGCTGCTGAGTTCTCCTGGCCACCATGAACTTCAGGA   50

20-3                  TAGCACCTGCCTGAACTACACAATCTGGGCTGTGGTGTA   39
nr4     AGTGGGTGCTATAGCAGCTGCCTGAACTACACAATCTGGGCTTTGGTGTA  100
nr61    AGTGGGTGCTATAGCAGCTGCCTGAACTACACAATCTGGGCTTTGGTGTA  100
9-2     AGTGGGTGCTATAGCAGCTGCCTGAACTACACAATCTGGGCTTTGGTGTA  100
4-4c    AGTGGRTGCTATAGCAGCTGCCTGAACTACACAATCTGGGCTTTGGTGTA  100
1-12d   AGTGGGTGCTATAGCAGCTGCCTGAACTACACAATCTGGGCTTTGGYGTG  100
CON     AGTGGGTGCTATAGCAGCTGCCTGAACTACACAATCTGGGCTTTGGTGTA  100

20-3    T.CCCTGTATGCCCTCCGGGCCAGACACTGGAGGTGTCATTTCCAAAGCA   89
nr4     T.CCCTGTATGCCCTCCGGGNCAGACACTGGAGGTGTCATTTCCAAAGCA  150
nr61    TACCCTGTATGCCCTCCGGGCCAGACACTGGAGGTGTCATTTCCAAANCA  150
9-2     T.CCCTGTATGCCCTCCGGGCCAGACACTGGAGGTGTCATTTCCAAAGCA  150
4-4c    T.CCCTGTATGTCCTCCGGGYCAGACACTGGAGGTGTCATTTCCAAAGCA  150
1-12d   T.CCCTGTATGCCCTCCGGGCCAGACACTGGAGGTGTCATTTCCAAAGCA  150
CON     T.CCCTGTATGCCCTCCGGGCCAGACACTGGAGGTGTCATTTCCAAAGCA  150

20-3    AATTGGAAGCGCTTTTTTGG.AATTTCTCTCCNATGCTTTCTACTCACAA  139
nr4     AATTGGAAGCG-----------------------------------------  200
nr61    AATTGGAAGCGCTTTTTTGG.AATTTCTCTCCAATGCTTTCTACTCACAA  200
9-2     AATTGGAAGCGCTTTTTTGGGAATTTCTCTCCAATGCTTTCTACTCACAA  200
4-4c    AMTCGYTAGCGCYCCCCCGG.AATHTCTCTCCAATGCTTTCTACTCACAA  200
1-12d   AATTGGAAGCGCTTTYTTGG.AATTTCTCTCCAATGCTTTCTACTCACAA  200
CON     AATTGGAAGCGCTTTTTTGG.AATTTCTCTCCAATGCTTTCTACTCACAA  200

20-3    .AGACTGACATCTTAACACGTGGCAAAGAAAAAAATATTTAAAGGGTCCA  189
nr4     ----------------------------------ATTTGGNGGGTCCA  250
nr61    .AGACTGACATCTTAACACGTGGCAAGGAANAAAATATTTAAAGGGTCCA  250
9-2     .AGACTGACATCTTAACACGTGGCAAAGAAAAAAATATTTAAAGGGTCCA  250
4-4c    CMGTCTGACATCTCATCACATGGTAAAGMACATATCATTTCTCT        244
1-12d   .AGACTGACATCTYAACACGYGGCAAAGAAAMMCMTMTTTAAAGGGTCCA  250
CON     .AGACTGACATCTTAACACGTGGCAAAGAAAAAAATATTTAAAGGGTCCA  250
```

```
20-3   GATC                                                          193
nr4    GATC.TATTTATGTAAACAATCAAGAGTGAGTTTGTAGTGGAAAACCCAA           300
nr61   GATC                                                          254
9-2    GATCTTWTTTATGTAAACAATCAAGAGTGAGTTTGGGGGTGG                   292
1-12d  GATC                                                          254
CON    GATCTTATTTATGTAAACAATCAAGAGTGAGTTTGKRGKKGRAAACCCAA           300

nr4    AGTTGGANAANTGGTGCATAATAAAAGAAATTTGTTCATTTTCTGCTGTT           350
CON    AGTTGGANAANTGGTGCATAATAAAAGAAATTTGTTCATTTTCTGCTGTT           350

nr4    GTACATTTGGAATGATTTTTGTATTTTGATTTTGTGAACATGTCTCCTAA           400
CON    GTACATTTGGAATGATTTTTGTATTTTGATTTTGTGAACATGTCTCCTAA           400

nr4    TGTAAATAGCCAATAGATTCTCTTTTCGGACAGTATCTTCTGATAGCTGG           450
CON    TGTAAATAGCCAATAGATTCTCTTTTCGGACAGTATCTTCTGATAGCTGG           450

nr4    AATGTCTGGGTTATAAAATTTGTGATC                                  477
CON    AATGTCTGGGTTATAAAATTTGTGATC                                  477
```

B) Family 2

'nr' = nonran, '15-3' = pr15-3.

```
2-10b GATCTGGTGGAAACACAGGCTGCCAGGCCCTGCCCCTGGAGTCTTTCTTT    50
nr63  GATCTAGTGGAAACACAGGCTGCCAGGCCCTGCCCCTGGAGTCTTTCTTT    50
4-7g  GATCTGGTGGAAACACAGCCTGCCAGCCCCTGCCCCTGGAGTCTTTCTTT    50
3-5g  GATCTGGTGGAAACACAGGCTGCCAGGCCCTGCCCCTGGAGTCTTTCTTT    50
CON   GATCTGGTGGAAACACAGGCTGCCAGGCCCTGCCCCTGGAGTCTTTCTTT    50

1-3c                                GATCAGAAGAAC.GGTTTCCTG    22
2-10b GAATAGGCCTGGGGTGGGGCCCAAGAATGAGCAGAAGAACAGGTTTCCTG   100
nr63  GAATAGGCCTGGGGTGGGGCCCAAGAATGAGCAGAAGAACAGGTTTCCTG   100
4-7g  GAATAGGCCTGGGGTGGGGCCCAAGAATGAGCAGAAGAACAGGTTTCCTG   100
3-5g  GAATAGGCCTGGGGTGGGGCCCAAGAATGAGCAGAAGAACAGGTTTCCTG   100
CON   GAATAGGCCTGGGGTGGGGCCCAAGAATGAGCAGAAGAACAGGTTTCCTG   100

15-3                              GATCCCACGTTGAGGACGGAGGGCT    25
4-7e                              GATCCCACGTTGAGGARGGAGGCCT    25
1-3c  GTGAGGCTGATGCCGCTGGCCCAGGGTTTTCACGTTGAGGACCGAGACCT    72
2-10b GTGAGGCTGATGCCGCTGGCCCAGGGTTCTCACGTTGAGGACCGAGAGCT   150
nr63  GTGAGGCTGATGCCGCTGGCCCAGGGTTCTCACGTTGAGGACCGAGAGCT   150
4-7g  GTGAGNCTGATACCGCTGGCCCAGGGTTCTCACGTTGAGGACCGAGAGCT   150
3-5g  GTGAGGCTGATGCCGCTGGCCCAGGGTTCTCACGTTGAGGGCCGAGAGCT   150
CON   CTGAGGCTGATGCCGCTGGCCCAGGGTTCTCACGTTGAGGACCGAGAGCT   150

15-3  TTGAAGGTTTTCATGCCTGATAATGGCCAGGAACCCTTCTCAGTAGGCAT    75
4-7e  TTGA.GKTTATCATGCCTGATGATGCCCAGGAACCCTTCTCAGTAGGCAK    75
1-3c  TTGA.GGTATTCATACCTGATGATGTCCAGGAACCCTTCTCAGTAGGCAC   122
2-10b TTGA.GGTTTTCATACCTGATGATGTCCAGGAACCCTTCTCAGTAGGCAC   200
```

```
nr63   TTGA.GGTTTTCGTACCTGATGATGTCCAGGAACCCTTCTCAGTAAGCAC  200
4-7g   TTGA.GGTTTTCATACCTGATGATGTCCAGGAACCCTTCTCAGTAGGCAC  200
3-5g   TTGA.GGTTTCCATACCTGATGATGTCCAGGAACCCTTCTCAGTAGGCAC  200
CON    TTGA.GGTTTTCATACCTGATGATGTCCAGGAACCCTTCTCAGTAGGCAC  200


15-3   TGAAGACCA...GCAGAGTCCCAGACCCCAGGAGAGATGTCGTCAGACGG  125
4-7e   CGAAGACCA...GCAGAGTCCCAGACCCCAGGAGAGATGTCGTCAGACGG  125
1-3c   TGAAGACCATCANCAGAATCACAGACCCCAGGAGAGATGTCNTCAGACAG  172
2-10b  TGAAGACCATCAGCAGAATCACAGACCCCAGGAGAGATGTCGTCAGACAG  250
nr63   TGAAGACCATCAGCAGAATCACAGACCCCAGGAGAGATGTCNTCAGACAG  250
4-7g   TGAAGACCATCAGCAGAATCACAGACCCCAGGAGAGATGTCGTCAGACAG  250
3-5g   TGAAGACCATCAGCAGAATCACAGACCCCAGGAGAGATGTCGTCAGACAG  250
CON    TGAAGACCATCAGCAGAATCACAGACCCCAGGAGAGATGTCGTCAGACAG  250


15-3   ACACAGAGGCATCACGGAATTAAAGTGAAAATGAAGAAAGGAGCT.GAGC  175
4-7e   ACACAGAGTCATCACGGAAKTAAAGTGAAAATGAAGAAAGGAGCT.GAGC  175
1-3c   ACGCAGAGGCATCACCGAATTAAAGTGAGAATGGAGAAAGGAGCT.GAGC  222
2-10b  ACACAGAGGCATCACCGAATTAANGTGAAAATGAAGAAAGGAGCTTGAGC  300
nr63   ACACAGAGNCATCACCGAAATAAAGTGTAAATGAAGGAAGGAANT.GAAC  300
4-7g   ACACAGAGGCATCACCGAATTAAAGTGAAAATGAAGAGAGGAGCT.GAGC  300
3-5g   ACACAGAGGCATCACCGACATTAACGTGANAATGAAGAAAGGAG        294
CON    ACACAGAGGCATCACCGAATTAAAGTGAAAATGAAGAAAGGAGCT.GAGC  300


15-3   ATCTGTTTCATGACTTC.CCTGCGTTATTTTA.CTAAAGAGGCTGTTCTG  225
4-7e   ATCTGTTTCATGACTTT.CCTGCGCTGTTTTA.CTAAAGAGGCCGTTCTG  225
1-3c   ATCTGTTTCATGACTTT.CGTG.GCTGTTTTA.CTAAAGAGGCTATCCTG  272
2-10b  ATCTGTTTCATGACTTTTCGT                               321
nr63   ATCTGTNTCATGACTTT.CGTG.GCTGTTTTATCTAGAGAGCCAATCCTG  350
4-7g   ATCTGTTTCATGACTTT.CGTG.NCTGTTTTA.CTAAAGAGGCTATCCTG  350
CON    ATCTGTTTCATGACTTT.CGTG.GCTGTTTTA.CTAAAGAGGCTATCCTG  350


nr2A                              ATCCCCCACCANNTGCTGNACNATGTCCT   29
15-3   GCCCAGTCAGGGCACGCTATCATCACCAACTACTTGTTGAACTATGTCCT  275
4-7e   GCCCAGTCAGGGCACGCTATCATCACCAACTGCTTGCTGAACTATGCCCT  275
1-3c   NCCCAGTCAGGGCACNCTATCATCNCCAACTACTTGTTGAACTACNTTC.  322
nr63   GCCCATTCAGGGCACGCTATCNTCCCCAACTACTTGTTGAACTACGTTCA  400
4-7g   GCCCAGTCAGGGCACGCTATCATCACCAACTACTTGTT               388
CON    GCCCAGTCAGGGCACGCTATCATCACCAACTACTTGTTGAACTATGTCCT  400


nr2A   GGNTNTTGACCCTGANGGATGGAATCCCTCGGG.TTGCAACCCACCATGC   79
15-3   GGGTCTTGACCTTGAAGGATGGATGC.GTTTGGGTTGCAAGCCGCCATGC  325
4-7e   GGGTCTTGACCTTGAAGGATGGACGC.GTTCGGGTTGAAAGCCGCCATGC  325
1-3c   AGGG.T...CTTGATA                                    338
nr63   GGNTCTTGATC                                         411
CON    GGGTCTTGACCTTGAAGGATGGAHGC.GTTBGGGTTGCAAGCCGCCATGC  450


nr2A   AGGGAGCGAACCGATC    95
15-3   AGGGTCGAACCGATC    340
4-7e   AGGGTCGAACCGATC    340
CON    AGGGTCGAACCGATCC   466
```

C) Family 3

```
4-7a   GATCCGTGGGAGCACTGAGATGGTCATCTCCAAATAT.CACATATCTCAC    50
5-3c   GATCCGTGGGAGCACTGAGATGGTCATCTCCAAATATCCACATATCTCAC    50
3-4e   GATCCGTGGGAGCACTGAGATGGTCATCTCCAAATAT.CACATATCTCAC    50
4-10a  GATCCGTGGGAGCACTGAGATGGTCATCTCCAAATAT.CACATATCTCAC    50
CON    GATCCGTGGGAGCACTGAGATGGTCATCTCCAAATAT.CACATATCTCAC    50

4-7a   ATCCTCAGAAAGGCTTATATAGCAAGGAATCATTTCCGTGCTCAGAGCAC   100
5-3c   ATCCTCAGAAAGGCTTATATAGCAAGGAATCATTTCCGTGCTCAGAGCAC   100
3-4e   ATCCTCAGAAAGGCTTATATAGCAAGGAATCATTTCCGTGCTCAGAGCAC   100
4-10a  ATCCTCAGAAAGGCTTATATGGCAAGGAATCATTTCCGTGCTCAGAGCAC   100
CON    ATCCTCAGAAAGGCTTATATAGCAAGGAATCATTTCCGTGCTCAGAGCAC   100

4-7a   TATCCACTCTGAAGTNGTTGAGCATGTCTAGGACTCCAGTTATGCACCAT   150
5-3c   TATCCACTCTGAAGTTGTTGAGCATGTCTAGGACTCCAGTTATGCACCAT   150
3-4e   TATCCACTCTGAAGTGGTTGAGCATGTCTAGGACTCCAGTTATGCACCAT   150
4-10a  TATCCACTCTGAAGTTGTTGAGCATGTCTAGGACTCCAGTTATGCACCAT   150
CON    TATCCACTCTGAAGTTGTTGAGCATGTCTAGGACTCCAGTTATGCACCAT   150

4-7a   GAAGTGAGCTCTGGGTTCACTGGCTGGGGCTTTTGGATCTGTGCCAA     197
5-3c   GAAGTGAGCTCTGGGTTCACTGGCTGGGGCTTTTGCATCTGTGCCAA     197
3-4e   GAAGTGAGCTCTGGGTTCACTGGCTGGGGCTTTTGGATCTGTGCCAAATC  200
4-10a  GAAGTGAGCTCTGGGTTCACTGGCTGGGGCTTTTGCATCTGTGCCAAATC  200
CON    GAAGTGAGCTCTGGGTTCACTGGCTGGGGCTTTTGSATCTGTGCCAAATC  200

3-4e   AGTCCTTGCTGATACATTTCTCACATCCTTG                     231
4-10a  AGTCCTTGCTGATACATTTCTCACATCCTTG                     231
CON    AGTCCTTGCTGATACATTTCTCACATCCTTG                     231
```

D) Family 4

```
4-7d   GATCCAAATGGCATCTCCTACCCTATCCCTGTATGGCTTAACCTAGGGAA   50
CON    GATCCAAATGGCATCTCCTACCCTATCCCTGTATGGCTTAACCTAGGGAA   50

4-7d   ACTCTTTCCCAGAGGAGTCAGGTAAGGAGACGGTGGCTGAGCTTCTTACA  100
CON    ACTCTTTCCCAGAGGAGTCAGGTAAGGAGACGGTGGCTGAGCTTCTTACA  100

5-7e                                GGGACACSAKCCTGCCATTCTGT.  24
4-7d   GACTTAAAGGAGGCATCCTGGAATTTAGGAGTCCGTCCTTCCTTTCTGTA  150
CON    GACTTAAAGGAGGCATCCTGGAATTTRGGASWCCRTCCTKCCWTTCTGTA  150

5-7e   .CCACTGGCAGCKGCTGCTGCTTCGCAGAG.TTAGSTTTGTCTCTCAACC   74
4-7d   ATCCCTGGCAGCTCCTGCTGCTGCTCAAAGTTTAGCTTTGTCTCTCATCC  200
CON    AYCMCTGGCAGCTSCTGCTGCTKCKCARAGTTTAGCTTTGTCTCTCAWCC  200

5-7e   AGCWYAGRCTGGTGCTGGTCCTGATGGCCTCTGCATAGCTGTATTAGKAT  124
4-7d   AGCTCAGACTGTTGCTGGTCCTGATGGCCTCTGCTTAGCTGTATTA....  250
CON    AGCTCAGACTGKTGCTGGTCCTGATGGCCTCTGCWTAGCTGTATTAG.AT  250
```

```
5-7e  GTKCTTGCRTTGCWAAAAAGAAATACATGAGARATACCTTTAGKTGGCTG  174
4-7d  GTC                                                 253
CON   GTCCTTGCRTTGCWAAAAAGAAATACATGAGARATACCTTTAGKTGGCTG  300

5-7e  ATGGTWCCATAGGGDATACAGGAAGCRTGGTGGTTDCTGCTTGTGGGGAG  224
CON   ATGGTWCCATAGGGDATACAGGAAGCRTGGTGGTTDCTGCTTGTGGGGAG  350

5-7e  GCCTCAGGAAACTTTTACTCAAGGTACAGGCCCAATGAGKAAAAAGAAAA  274
CON   GCCTCAGGAAACTTTTACTCAAGGTACAGGCCCAATGAGKAAAAAGAAAA  400

5-7e  AAAAAAAAGCTACCAGCRCCGCAATGTGGAGTCCTGAKC              313
CON   AAAAAAAAGCTACCAGCRCCGCAATGTGGAGTCCTGAKC              439
```

E) Family 5

```
2-5c  GAATTCTATGCCTCAGGTTYTCATTTTTGTTATGTGCAAAATAAACTCGT   50
CON   GAATTCTATGCCTCAGGTTYTCATTTTTGTTATGTGCAAAATAAACTCGT   50

2-5c  ATCCTGCTATAGATYTAAACCTGTGAGVGAMTCCTACTCAAAACCTGATT  100
CON   ATCCTGCTATAGATYTAAACCTGTGAGVGAMTCCTACTCAAAACCTGATT  100

2-5c  CAAACTTGTGTGTAGATCTTTGTCTCTCTGTGGCATACAGAATCAATKCT  150
1-11d                 GATCTNTGCNTCTCTGTGGCATAAGTAATAATTCCT   36
CON   CAAACTTGTGTGTAGATCTTTGYCTCTCTGTGGCATAMRKAATMAWTCCT  150

2-5c  TCTGGYTYACTTCCCTCAGAAMAATGGACTTAGACTTCCCWCAAGACTTC  200
1-11d TCTGGTTTCCTTCCCTCAGAATAATGGANTTAGACTGCCCACAAGCCTTC   86
CON   TCTGGTTTMCTTCCCTCAGAATAATGGACTTAGACTKCCCACAAGMCTTC  200

2-5c  CAGVTAGTCACMACCTGTCTTATCTGCCTRAGTTACCTCATAGACCCCAT  250
1-11d CGGACAGAACTCACCTGGCTTATCTGGCTGAATTACCTCATAGACCCCAT  136
CON   CRGAYAGWMMYCACCTGKCTTATCTGSCTGARTTACCTCATAGACCCCAT  250

2-5c  CACTATAGGCTGGGGGSGACAGCCTCTGTAGGCCCTGCCTCTSSCYTCGCT  300
1-11d CACTATAGGCTGCGGGNACAGNTTCTGTAGGCCCTGCCTCTGCCTTAGCT  186
CON   CACTATAGGCTGSGGGGACAGCYTCTGTAGGCCCTGCCTCTGCCTTMGCT  300

2-5c  GGGAAGAAGTACACACTCCTGCCCCGCATGCAGGGAATCGTCACAGCAGG  350
1-11d GGGAAGAAGCACACACTCCTGCCCTGCATGCA                    218
CON   GGGAAGAAGYACACACTCCTGCCCYGCATGCAGGGAATCGTCACAGCAGG  350

2-5c  AAGATTTC  358
CON   AAGATTTC  358
```

F) Family 6 (clone 4-1f only)

```
CGTGAGCCAC CACCCCCGGC CCCAGGACAC ACAGCTTTAA AATTTCTCCT   50
TGGTCTCACC CAGTGCCAAC CACCTAAAAC CTCTCATTTT CCCCCAGACA  100
TTTCTTCTGC CTCCAGGATG GAGGTAGAGA ATCTTGGCCT AGGCCCACGC  150
ACTGGGGACC ATGCTGGGCT GCCGTGGACA GTGACGGACT CAGGTTCTCA  200
CCAGGATC                                                208
```

G) Family 7 (clone 3-3a only)

```
GATCTAGGAG CAGAGGGCAG AGCCTCAGCA GGAAGAGCGT CTCTTTGAGA   50
AGGAGACACA GTGGAGCAGG TGTGTAGGTT CACAGGGCCA GCTATGGGTA  100
GAGTCGGGTG TACATTTTTA GGAGCCACAA TTCCCAAAAA TCTCCTGACT  150
ATAACATCAG TGCACAGAGC CAGTCAAATG GAGGAGGAGT GGGTCCAGGC  200
AATTCAGGAA GAAGGAAAGT AACAAATGAG TGGTTGCAGG AGGACACTTT  250
TTCTTGTCGA GGGTCACTAA ACAAAACATT GTCTCCTCCC CCTCTAACTT  300
CAGAAACAAT GGAGGGTAAG AGTGTCNCCT GGGCCCTGGG GCCAAAGACA  350
GTAGATAACT TCTCTGTCGT GTTCTCCAGA AGGGCCCAAC AANTACAAGG  400
TTCTACGGTT CTAAATTCCA ATCTAGTCTT CCACATCATT TTGAAGGTAT  450
AATATTACTT GTCAAAGTGG GATGATAGAA GATATGTGTG GACATAAATT  500
GTTGTCAA                                                508
```

CHAPTER 5

ANALYSIS OF THE D0485 ALPHA TUBULIN SEQUENCE

The large number of alpha tubulin-related clones identified from the two CSC product libraries was very strongly suggestive of a tubulin-related sequence located within the YAC D0485. Before any further analysis however, it was necessary to prove this by hybridisation. Secondly, due to the high degree of homology between the coding sequences of the known α-tubulin subtypes it was impossible to deduce whether the D0485 α-tubulin gene had the same sequence as any of the known subtypes, or whether it represented a novel subtype. Detailed sequence analysis was required to investigate the latter.

5.1 Map Position of the D0485 Tubulin Gene

The position of the translocation breakpoint on YAC D0485 had recently been identified by detection of an altered restriction fragment size between normal and translocation genomic DNA, and by FISH analysis of cosmids which had been placed on the long range restriction map of the YAC. It was thus possible to use three CSC α-tubulin clone probes both to prove the existence of an α-tubulin gene within D0485 and to investigate its location relative to the translocation breakpoint. The clones chosen were 2-2b, random 4 and 1-6h. 2-2b is located at the 5' end of the gene, random 4 is centrally located and 1-6h is at the 5' end of the gene. Both 2-2b and 1-6h contain sequence from the UTRs (which are not conserved between subtypes), yet both contain sufficient of the coding sequence to enable hybridisation to all subtypes.

The three probes were hybridised to *Not I*, *Sal I* and *Sfi I* restriction fragments of YAC D0485 resolved by PFGE. Figure 5.1 shows the result obtained with probe 2-2b. It can be seen that the α-tubulin probe hybridised to a single band in each track, of size 1Mb (*Not I*), 350kb (*Sal I*) and 150kb

(*Sfi I*). The other probes random 4 and 1-6h also hybridised to these same bands. This confirmed that there was indeed at least one α-tubulin gene on YAC D0485, and also enabled placement on the pulsed field map of the YAC (Figure 5.2). The α-tubulin gene was situated within the same *Not I* and *Sal I* restriction fragment as the translocation breakpoint. Hybridisation onto a set of the somatic cell hybrids (by Sheila Christie) revealed that the α-tubulin locus lay centromeric to the chromosome 11 breakpoint.

The same three probes were also used to isolate α-tubulin positive cosmids from a library made from YAC D0485 by Yoshiro Shibasaki. FISH analysis of these cosmids on chromosomes from a translocation carrier (by Yoshiro Shibasaki) confirmed the location of α-tubulin centromeric to the breakpoint and estimated the distance from α-tubulin to the breakpoint as approximately 250kb ± 40kb. Fingerprint analysis of the cosmids and construction of a contig from them (by Sheila Christie and Yoshiro Shibasaki) provided no evidence to suggest that there was more than one copy of the gene. One end of the contig was fixed by the presence of a *Sal I* site in several of the cosmids (see figure 5.2). This enabled the more accurate estimation of distance from the α-tubulin locus to the breakpoint as approximately 300kb.

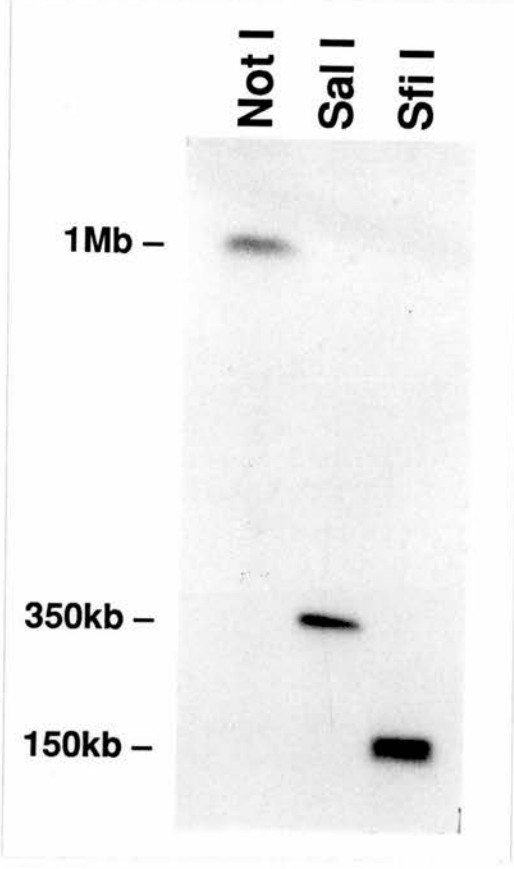
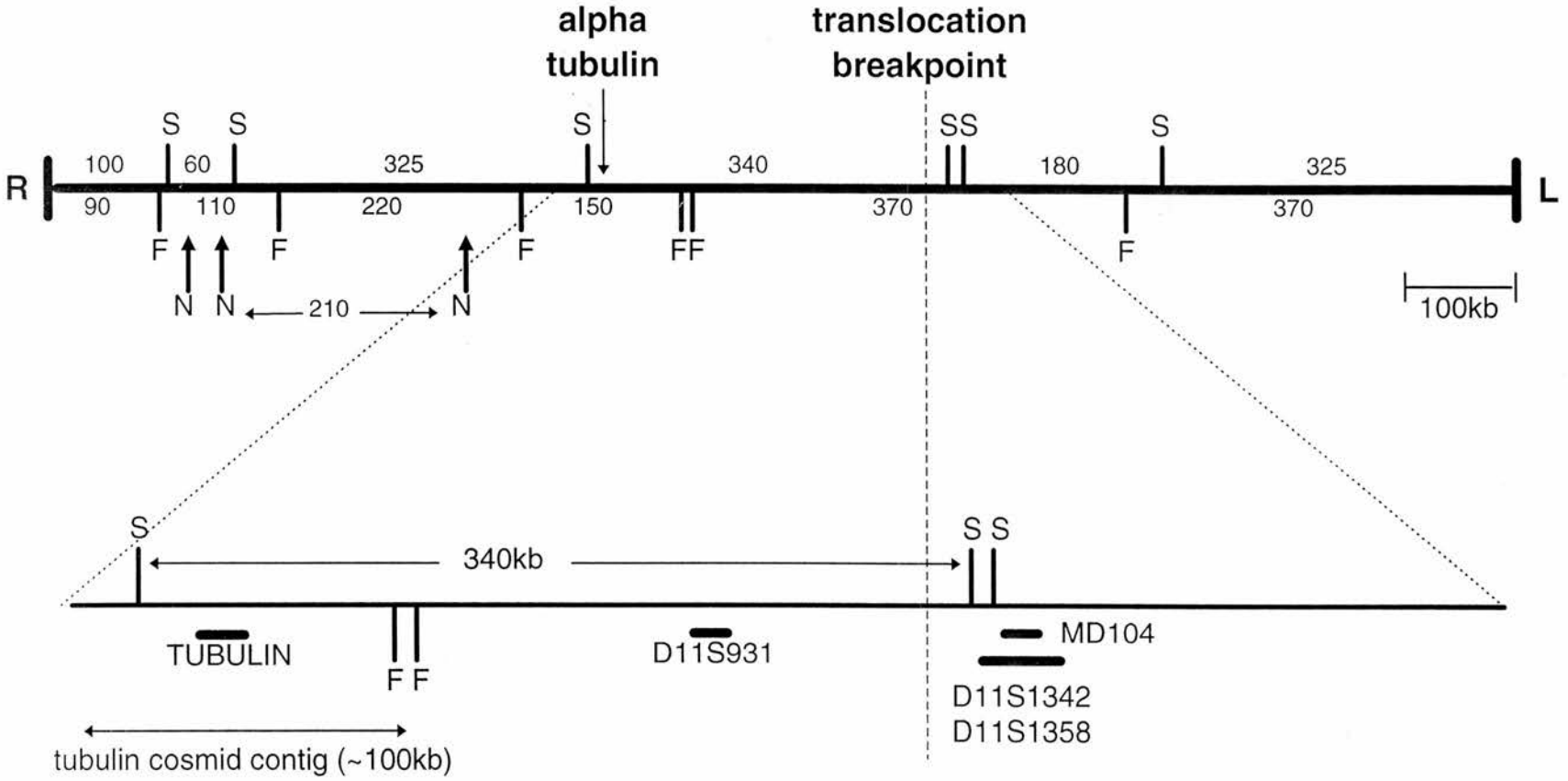


Figure 5.1 Hybridisation of the CSC α-tubulin clone probe 2-2b to restriction digested fragments of YAC D0485, resolved by PFGE. Track 1 contains a *Not I* digest, track 2 a *Sal I* digest and track 3 an *Sfi I* digest. The probe hybridises to a single band in each track, of size 1Mb (*Not I*), 350kb (*Sal I*) and 150kb (*Sfi I*).

Figure 5.2 *Sal I* (S), *Sfi I* (F) and *Not I* (N) restriction map of YAC D0485 showing the positions of the translocation breakpoint and the α-tubulin-related locus. The distances between the restriction sites are shown in kilobases. The enlarged region (not to scale) also shows the positions of the genetic markers D11S931, D11S1342 and D11S1358 plus the microdissection clone marker MD104. (Pulsed field mapping of this YAC was done by John Brown, John Maule and myself.)



5.2 Sequence of the D0485 Alpha Tubulin Gene

5.2.1 Sequencing of cosmid subclones

To enable the sequencing of the α-tubulin gene from an α-tubulin positive cosmid, it was first necessary to subclone the cosmid into a series of clones of more useful size, and then select those which were positive for α-tubulin. Two of the α-tubulin positive cosmids, 39 and 453, were subcloned by Tony Brookes. The cosmid was digested with *Sau3AI* to varying degrees of complete digestion (very partial, partial and complete) by the production of an enzyme dilution series. 8ng of digested cosmid DNA was then ligated to 10ng pBS vector which had been linearised with *BamHI* and the cut ends phosphatased. Transformation of this ligation mix resulted in recombinant clones.

In order to select for fragments of the cosmid which contained parts of the α-tubulin gene, filters lifted from the subcloned library were hybridised with the 5' α-tubulin clone probe 2-2b and an oligonucleotide probe, 1000 (see table 2.2) derived from the 3' UTR sequence in the clone 1-6h (the latter by Tony Brookes). Hybridisation with 2-2b resulted in five definite positives (termed 5'1 to 5'5). These colonies were picked, and their inserts amplified by PCR (see table 5.1). Clones 5'1, 5'2 and 5'5 were derived from the 'partial digest' library, clones 5'3 and 5'4 were picked from the 'very partial digest' library. As expected, the clone with the smallest insert size (5'1) was derived from the less partial of the two libraries and the clone with the largest insert size (5'4) was derived from the more partial library. Hybridisation of the oligonucleotide 1000 was performed at 50°C, then the filters were washed at 50°C in 4x SSC, 0.1% SDS. Five positively hybridising colonies were picked (termed 3'1 to 3'5). Clone 3'1 was the same as clone 5'4. PCR amplification of the inserts of 3'2 and 3'4 enabled estimation of the clones' insert sizes (see table 5.1).

Table 5.1	
Clone	**Estimated insert size (kb)**
5'1	0.8
5'2	1.6
5'3	1.2
5'4	2.5
5'5	1.5
3'2	1.3
3'4	1.3

The ends of clones 5'2 to 5'5, 3'2 and 3'4 were sequenced by ABI automated cycle sequencing, using the primers 291 and 292 derived from the ends of the pBS sequence. This enabled alignment of these six clones relative to the known α-tubulin sequences (see Figure 5.3).

The 300-400bp of sequence data from each end of the six subclones was insufficient to cover the entire α-tubulin gene. In addition, sequence data was required from both DNA strands to resolve ambiguous sequences or correct inaccuracies obtained from one strand. Hence more primers were designed from the new sequence as it became available. These primers were 309, 312, 313, 417, 467, 468, 469, 471, 709 and 710 (see table 2.2). Sequence data was collated by the GCG fragment assembly computer programs. Multiple sequencing runs using these primers on miniprep DNA derived from an appropriate subclone enabled a consensus sequence of 2683bp to be assembled. The sequence data extended considerably further than the α-tubulin gene sequence itself, 625bp upstream and 479bp downstream of the published keratinocyte α-tubulin sequence. The entire consensus had been sequenced at least twice in one direction and once in the other, apart from

the first 278bp of sequence which was sequenced many times but only in one direction.

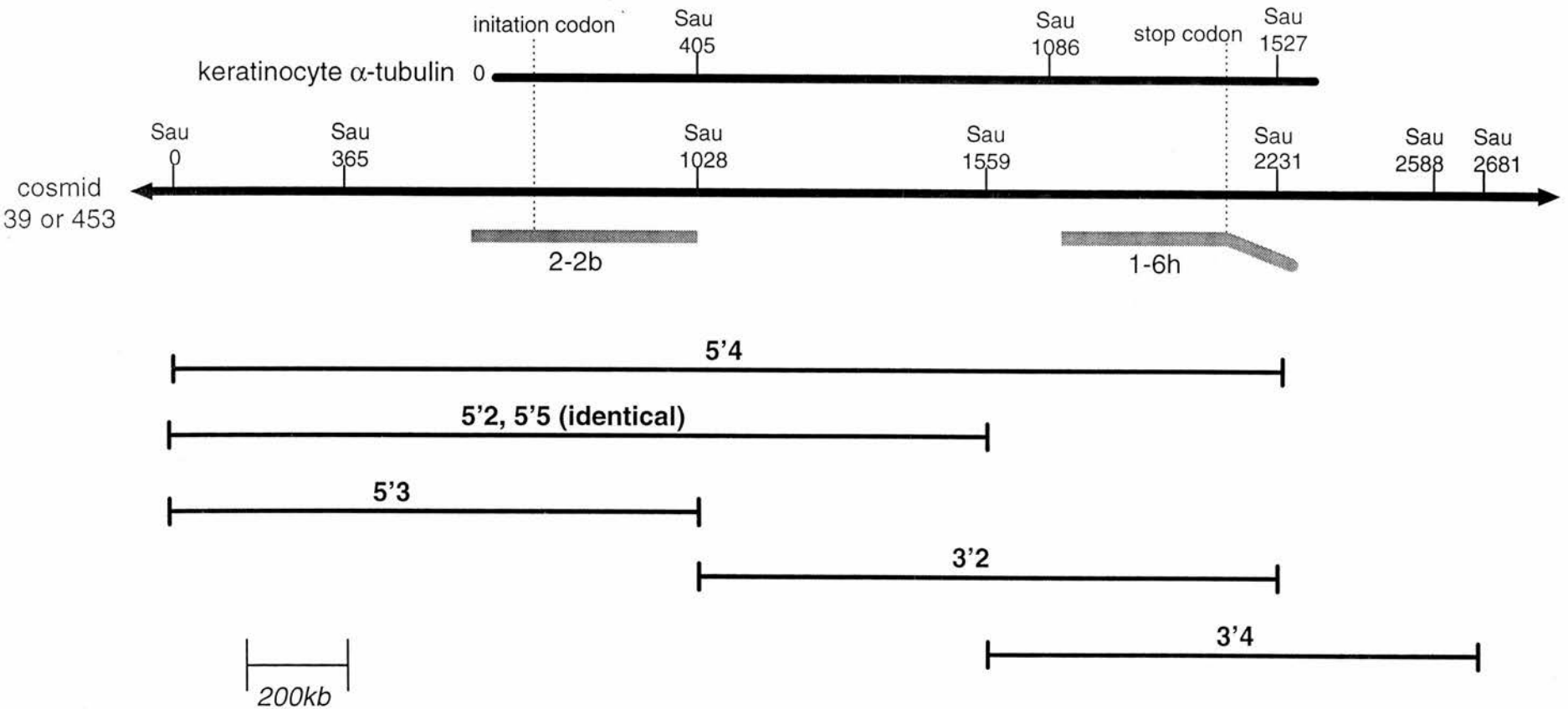
The sequence of the D0485 α-tubulin gene in relation to the previously known α-tubulin genes will be discussed in section 5.2.4.

5.2.2 Inaccuracy in the consensus sequence

As sequencing by ABI cycle sequencing is only 98-99% accurate, the derived consensus sequence was probably imperfect, despite multiple rounds of sequencing. In addition, the DNA which was sequenced had been subjected to three rounds of cloning steps - from genomic DNA firstly to a YAC, then to a cosmid and finally to plasmid subclones. Single base pair changes in sequence may occur during cloning, and since the DNA used in the above sequencing reactions was derived from clonal sources, a change of sequence would not be identified.

The consensus sequence also contained some ambiguities which could not be resolved by further rounds of sequencing. In these instances, one particular base was always obtained when the sequencing reaction was in one direction, and another base was always obtained from sequencing in the other direction. IUB ambiguity codes were therefore inserted into the consensus sequence.

Figure 5.3 Position of α-tubulin-selected subclones from cosmids 39 and 453 in relation to the cosmid and the keratinocyte α-tubulin sequence. *Sau3AI* sites are marked by 'Sau' and their numerical position in the sequence. The *Sau3AI* site in keratinocyte α-tubulin sequence at position 1086 is not present in the corresponding D0485 α-tubulin sequence. Likewise the *Sau3AI* site in the D0485 α-tubulin sequence at position 1559 is not present in the keratinocyte α-tubulin sequence. The position of the CSC clone probes 2-2b and 1-6h, used to isolate the subclones, is also shown. The 3' end of 1-6h is shown tilted since it is derived from foetal brain α-tubulin and therefore does not hybridise to the keratinocyte or D0485 cosmid α-tubulin genes.



5.2.3 Sequencing of the D0485 tubulin gene from a normal individual and a translocation carrier

Due to the possible inaccuracies in the α-tubulin consensus sequence obtained by the automated sequencing of clones, it was necessary to use a more accurate method to derive the genuine D0485 α-tubulin sequence from a normal individual. It was also important to ascertain whether this sequence remained the same in individuals carrying the t(1;11) translocation. Although the α-tubulin sequence had been mapped to 340kb away from the breakpoint, the possibility of a long range effect of the translocation could not be ruled out. The translocation may remove or insert new enhancer elements necessary for transcription, or it may introduce a different kind of chromatin environment. Sequencing reactions which employ PCR-amplified material as a DNA template rather than plasmid DNA are advantageous in cases where accuracy of sequence is important. Although errors of sequence will occur during amplification (the error rate of *Taq polymerase* is 1 error in every $25x10^{-6}$ bp; this enzyme was used although polymerases with lower misincorporation rates are available), an average of the millions of template molecules available for sequencing will produce an accurate sequence overall.

Due to the high similarity in the sequence of the known α-tubulin genes, it was important to ensure that the DNA amplified and sequenced was indeed from the D0485 α-tubulin gene alone. For this reason, primers 711 and 712 were designed (see table 2.2). These primers lie outside the α-tubulin sequence at the extremities of the consensus sequence derived from the ABI sequencing. These extremities are anonymous regions of chromosome 11 and therefore their sequence should be unrelated to that adjacent to other α-tubulin genes.

300ng each of primers 711 and 712 were employed to amplify a region containing the D0485 α-tubulin gene from 100ng of genomic DNA. The PCR programme used was:

Hot start at 90°C.

Denaturation: 94°C for 30 secs in the first cycle and 15 secs thereafter.

Annealing: Touch down from 66°C to 56°C then 56°C, for 30 secs.

Extension: 72°C for 1 min (cycles 1-10), 2 mins (11-20) and 4 mins (21-30).

Figure 5.4 shown the results obtained from this PCR amplification. A strong band of the expected size (2.7kb) was obtained from total human genomic DNA from a normal individual. Weaker bands of the same size were obtained from the cell line MAFLI (derived from a translocation carrier) and from the somatic cell hybrid MAR1 (in which the only chromosome 11 component is that from the der(11) from a translocation carrier). As further proof of the map location of the D0485 α-tubulin gene, no PCR product was obtained from MAR12, which contains the der(1) translocation chromosome and no chromosome 11 sequence from above the breakpoint. No product was obtained from total hamster or total mouse genomic DNA, proving that the band in MAR1 was genuinely derived from the human component of the hybrid.

It was necessary to use MAR1 as the template from which to derive the sequence of the D0485 α-tubulin gene from a translocation carrier. MAFLI, although carrying translocation chromosomes, also contains a set of normal chromosomes. Unfortunately a strong, clean band could not be obtained from MAR1 by PCR amplification with primers 711 and 712, despite variation in the amount of template DNA used and the annealing temperatures of the PCR programme. Hence a secondary PCR was employed. To facilitate the production of a large amount of template for sequencing, the size of the target product was decreased by amplifying the tubulin sequence in smaller sections, initially two, using either 711 and 312 or 712 plus 469 (see table 2.2). There is only 7bp between the end of 312 and the beginning of 469, so it was still possible to sequence the vast majority of the α-tubulin gene.

Secondary PCR in two sections was also employed for the DNA from a normal individual, partly to increase the amount of template DNA available for the many sequencing reactions required, and partly because superior sequencing results were obtained from the shorter templates.

For the secondary PCRs, 1μl of primary (711-712) PCR product was reamplified with 300ng each of either primers 711 and 312, or 712 and 469. The PCR programme used was:
Hot start at 90°C
Denaturation: 94°C for 30 secs in the first cycle and 15 secs thereafter
Annealing: Touch down from 60°C to 50°C then 50°C, for 30 secs
Extension: 72°C for 1 min (cycles 1-10), 1.5 mins (11-20) and 2 mins (21-30)
This resulted in the production of a strong band in all cases which was then prepared for sequencing.

As much as possible of the α-tubulin gene was sequenced using the existing primers using a radioactive dideoxy terminator method. Where the sequence agreed with that obtained from the cycle sequencing, it was not repeated. Regions in which there were differences from the cycle sequencing consensus were sequenced at least twice. Since this method of sequencing only resulted in a maximum of 200bp of sequence per reaction, sequence from the existing primers alone did not cover the entire gene. The oligonucleotides 150 to 159 were thus designed to fill in the gaps (see table 2.2). These new oligonucleotides were used both as primers for sequencing reactions and some as primers for secondary amplification of the 711-712 product. The amplification conditions used were exactly the same as those given for the secondary PCR above. PCR products were obtained from amplification with the primer pairs 152-312, 153-151, 155-158 and 156-158 (see Figure 5.5). The 153-151 template spanned the short gap between 312 and 469, and hence the entire gene was covered by PCR product template DNA for sequencing.

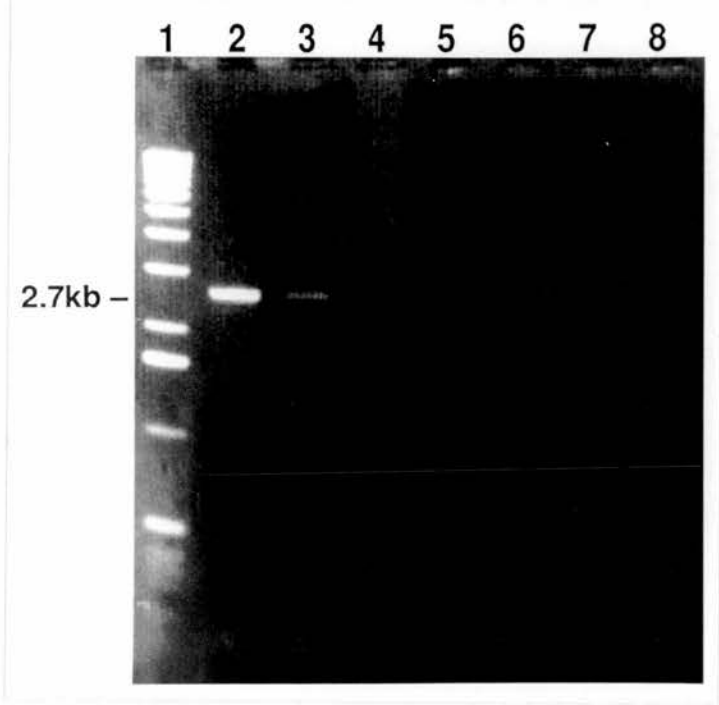


Figure 5.4 Amplification of the D0485 α-tubulin gene from genomic DNA.
The size marker used is 250ng of the 1kb ladder (track 1). Tracks 2 to 8 contain 10μl of the PCR product generated from different template DNAs with primers 711 and 712, resolved on a 1% agarose gel. Track 2 = total human genomic DNA, track 3 = MAFLI, track 4 = MAR1, track 5 = MAR12, track 6 = total hamster DNA, track 7 = total mouse DNA, track 8 = no DNA PCR negative control. A band of the expected size (~2.7kb) was obtained from total human genomic DNA, MAFLI and MAR1. This confirms that the PCR was specific for a human gene and that α-tubulin is located proximal to the translocation breakpoint.

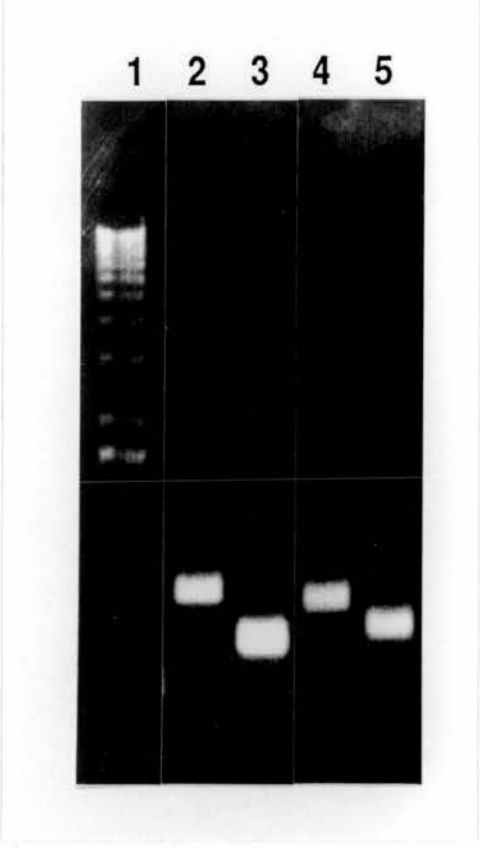


Figure 5.5 Amplification of fragments of the D0485 α-tubulin gene by nested PCR. Allele specificity was provided by the primary PCR. The size marker used is 250ng of the 1kb ladder (track 1). Tracks 2 to 5 show 10µl of the PCR product generated by re-amplification the primary PCR product with different primer pairs, resolved on a 1% agarose gel. The products generated were approximately 700bp with primers 152 and 312 (track 2), 500bp with primers 153 and 151 (track 3), 650bp with primers 155 and 158 (track 4) and 550bp with primers 156 and 158 (track 5).

Use of appropriate combinations of primers and template enabled conclusive sequence data to be obtained for the entire coding sequence of the D0485 α-tubulin gene, plus 450bp upstream and 160bp downstream, for both a normal individual and a translocation carrier (see Figure 5.6). In all cases, ambiguous sequence resulting from the cycle sequencing was resolved. There was however a small proportion of bases whose identity could not be determined due to signal across all four tracks ('stops') on the sequencing autoradiograph. These amounted to 1.9% (39 out of 1960 bases) for the 'normal' sequence and 1.8% (36 out of 1950 bases sequenced) for the 'translocation' sequence. In most cases, these stops occurred in the same places for each, and were present each time the region was sequenced. There were no positions in which a stop from the radioactive sequencing coincided with an ambiguous base from the cycle sequencing, so it was deemed unnecessary to attempt to resolve them. Only at one position did the cycle sequencing consensus differ from the radioactive sequencing consensus. This was the insertion of a guanine residue at position 313. This residue was clearly absent from multiple radioactive sequencing runs and hence has not been included in the D0485 α-tubulin final consensus sequence. This insertion may have been either an artefact of cloning, or a cycle sequencing error.

There were no differences in sequence identified between the normal individual and the translocation carrier. This implies that the translocation has not, by chance, fixed a variant or mutant allele of α-tubulin.

5.2.4 Sequence of the D0485 α-tubulin gene in relation to the keratinocyte α-tubulin gene

The most striking feature of the D0485 α-tubulin gene is that it is extremely similar to the keratinocyte α-tubulin cDNA sequence in the database, and as such it possesses no introns in genomic DNA. There are 77 positions where a difference is seen between the D0485 α-tubulin sequence and the 1598

bases of the keratinocyte α-tubulin sequence (see Figure 5.7). These differences are summarised in Table 5.2.

The majority of single base substitutions were purine to purine (40.4%) or pyrimidine to pyrimidine (36.8%). Only 10.5% of changes were pyrimidine to purine, and 12.3% were purine to pyrimidine. Both of the single base deletions, both of the insertions, two of the multiple deletions (6 bases altered) and 11 of the single base changes occurred in the UTRs. This corresponds to 27% of the total number of changes, even though the UTRs comprise only 15.3% of the total length of sequence. Since the coding sequence therefore contains fewer changes than would be expected by length alone, this may imply an evolutionary selective pressure imposed on it by functional constraint.

The predicted protein sequence transcribed from the D0485 α-tubulin gene is shown in Figure 5.8. Twenty four (51%) of the 47 single base substitutions occurring within the coding sequence are in the third position of the codon, only one of which would change the encoded amino acid (a conservative change from glutamic acid to aspartic acid at position 290 in Figure 5.8). In addition 74% of the first or second position substitutions result in a conservative change. The 3bp deletion of CTT, although out of frame itself, results in a deletion of a phenylalanine residue, and no change of reading frame. The seven base pair deletion in the D0485 α-tubulin sequence results in a shift of reading frame, which produces a change in the next two out of three amino acids before a termination codon is introduced, terminating the putative peptide.

Figure 5.9 shows a comparison of the D0485 α-tubulin predicted peptide sequence and the keratinocyte α-tubulin peptide sequence. The predicted protein derived from the D0485 α-tubulin sequence is 374 amino acids long, compared with the keratinocyte α-tubulin protein which is 451 amino acids long.

Table 5.2

Class of mutation with respect to the keratinocyte α-tubulin gene	Change	Frequency	Comments
Single base deletion	C G	1 1	Both in 5' UTR
Multiple base deletion	CTT GAGCTGT ATT TTG	1 1 1 1	Deletion of phenylalanine Shifts reading frame Across kerat. stop codon In keratinocyte 3' UTR
Insertion	A C	1 1	Both in 5' UTR
Single base substitution	T to C A to G G to A G to C C to G C to T A to C T to A A to T C to A	18 13 10 4 4 3 2 1 1 1	2 in the 69 bp of 5' UTR (2.9%) 47 in the 1353 bp of coding sequence (3.5%) 9 in the 176 bp of 3' UTR (5.1%)

Figure 5.6 Sequence of the alpha tubulin gene and the adjacent sequences from YAC D0485. All has been sequenced by automated ABI cycle sequencing. The sequence of bases 252 to 2213 has been confirmed by radioactive sequencing of PCR products. Primers used for PCR and/or sequencing are highlighted. The name and direction of the primer is underneath the highlighted region.

```
1     GATCTGTGTT CTTCAGAAGT GTCAAAGTCA TGAAAGACAA GGAAAAACTG
51    AGGAGCTGTC ACAGGTGGGA GGGGATTAAA GAGATATAAC AACTAAATTC
                                  711→
101   ACTGTGGTAC AGGCATACTT TGAAGATATT GTGGGTTTGG TTCTAGACCA
151   CTGCAATAAA GCAAATTTCA CAATAAAGTG AGTCATATAA TATTTTAGGT
201   TTCCCAGTGC ATGTAAAAGT CATAACCTAT TAAGTGTGCA ATAGCATGAT
            309 →
251   ATTTAAAAAA TCATATATAT CATATTTACA AATACTGTAT TGCTAAAAAT
301   GTTAATGATT AGTTGAGCCT TTACTGATGG AAAGGCTTGC CTCAATGTTG
                    ← 710
351   ATGGCTACTG ATTGATCAGG GTGGTGATTG CTGAAGGTTG GAGTGGCTGT
                                        709→
401   GGCAATTTTT AACAAGACAA CAGTGAAGTT TGCCTCATTG ACTGATTCTT
451   CCTTTCATGA AAAATTTCTC TGTTGCATGT AATGCTGTTC GATAGCATTT
                                          ← 467
501   TACCTAAGTA GGCCTTTTTT TTTTTTTTTT CCAAATTGGA GTCAATCCTC
551   TCAAACTCTG CTGCTGCTTT ATCAACTAAG TTTATGGAAC ATTCCCAGTG
601   TGTTACTTAC CTCAACTCTT AGCTTGTTGG GGACGGTAAC CGGGACCCAG
                  152 →
651   TGTCTGCTCC TGTCACCTTC GCCTCCTAAT CCCTAGCCAC TATGCGTGAA
701   TGCGTCTCCA TCCACGTTGG CCAGGCTGGT GTCCAGATTG GCAATGTCTG
        468→ ←417
751   CTGGGAGCTC TACTGCCTGG AACATGGCAT CCAGCCCGAT GGCCAGATGC
801   CAAGTGACAA GACCATTGGG GGAGGAGATG ACTCCTTCAA CACCTTCAGT
            150 →                          496→
851   GAGACGGGTG CTGGCAAGCA TGTGCCCCGG GCTGTGTTTG TAGACTTGGA
901   ACCCATGGTC ATTGATGAAG TTTGCACTGG CACCTACCGC CAGCTCTTCC
                                    ← 714
951   ACCCTGAGCA GCTCATCACA GGCAAGGAAG ATGCTGCCAA TAACTATGCC
                          153→
1001  CGAGGGCACT ACACCATTGG CAAGGAGATC ATTGACCTTG TGTTCGACCG
1051  AATTCGCAAG CTGGCTAACC AGTGCACTGG TTTTCAGGGC TTCTTGGTTT
                                          154→
1101  TCCACAGCTT TGGTGGGGGA ACTGGTTCTG GGTTCACCTC CCTGCTCATA
1151  GAACGTCTCT TAGTTGATTA TGGCAAGAAG TCCAAGCTGG AGTTCTCCAA
```

1201 TTACCCAGCG CCCCAGGTTT CCACAGCTGT AGTTGAGCCC TACAACTCCA

1251 TCCTCACCAC CCACACCACC CTGGAGCACT CTGATTGTGC CTTCATGGTA

← 312

1301 GACAATGAGG CCATCTGTGA CATCTGTTGT AGAAACCTCA ATATCGAGCG

469 →

1351 CCCAACCTAC ACTAACCTTA ACCACCTTAT TAGCCAGATT GTGTCCTCTA

155 →

1401 TCACTGCTTC CCTGAGATTT GATGGAGCCC TGAATGTTGA CCTGACAGAA

← 151

1451 TTCCAGACCA ACCTGGTACC CTACTCCCAC ATCCACTTCC CTCTGGCGAC

156 →

1501 ATATGCCCCT GTCATTTCTG CTGAGAAAGC CTACCATGAA CAGCTTTCTG

1551 TAGCAGAGAT CACCAATGCT TCCTTTGAGC CAGCCAACCA GATGGTGAAA

1601 TGTGACCCTC GCCATGGTAA ATACATGGCT TGCTGCCTAT TGTACCATGG

1651 TGATGTGGTT CCCAAAGATG CCAATGCTGC CATTGCCACC ATCAAGACCA

1701 AGCGTAGCAT CCAGTTTGTA GATTGGTGCC CCACTGGCTT CAAGGTTGGC

715 →

1751 ATCAACTACC AGTCTCCCAC TGTGGTGCCT GGTGGAGACC TGGCCATGGT

1801 ACAGAGTGCA TGCTGAGCAA CACCACAGCC ATTGCTGAGG CCTGGGCTCG

← 497

1851 CCTGGACCAC AAGTTTGACC TGATGTATGC CAAGCGTGCC TTTGTTCACT

← 471

1901 GGTACGTGGG TGAGGGGATG GAGGAAGGCG AGTTTTCAGA GGCCCGTGAA

1951 GACATGGCTG CCCTTGAGA AGGATTATGAG GAGGTTGGT GTGGATTCTGT

← 158

2001 TGAAGGAGAG GGTGAGGAAG AAAGAGAGGA ATACTAATCC ATTCCTTTTG

159 →

2051 GCCCTGCAGC ATGTCGTGCT CCCAGAATTT CAGCTTCAGC TTAGCTGACA

2101 GACATTAAAG CTTTCTGGTT AGATTTTCAT TTGGTGATCA TGTCTTTTCC

313 → ← 157

2151 ATGTGTACCT GTAATGTTTT TCCATCATGT CTCAAAGTAC AGTCATTAAC

2201 ATAAAAAAAA AAGTTTATGT TATATTTTAG ATCATGTGGT GGCATTTCAA

2251 CAATGTTCAC AGYATCTTCA CTAAGAGTCG ATTGGGTCTC AAGAAMCCAC

2301 TTTTTTCTCC CATCCAGAGG AAGAAGCTTC TCATCCACTC AAGTYTCATT

2351 ATGAGATTGS AGCAATTCAG TCACATCTTC AGGCTCCACT TCTAATTGTA

2401 GCTCACTTGC TATTTCCACA TCTGCAGTTA CTTCCTCTAA TAATGTCTTG

2451 AATACTCAAA GTCATCAACT TCTTCAAACT CCTGTTATGG TTGACATTTT

2501 GACCTTCTCT CATAAATCAC TAATGCTCTT AATGACATCT TAGAATGGTT

2551 AATTATTTCC CAAAGTTTTT CAACTTGTTT AGTCAAGATC CATCAGAAGA

2601 TTCACAATTT ATATGGAAGC TATAGCCTTA TGAAATGTAC TGCTGAACAA

← 712

2651 ATGAGACTTG AAAATTGAAA TTTCTCTTTG ATC

Figure 5.7 Comparison of the sequence of the D0485 α-tubulin gene and adjacent non-coding sequence ('D0485', upper line) with the published cDNA sequence of the keratinocyte α-tubulin gene ('kerat', lower line). The keratinocyte sequence begins at D0485 sequence position 625 and ends at position 2204. Bases identical between the two are shown by " - ". Deletions are shown by " * ". The initiator codon ATG is found at position 692. The codon for termination of transcription of the keratinocyte gene is at position 2033 (TAA). In the D0485 gene however a seven base deletion in the coding sequence results in premature termination of transcription at the TGA at position 1814.

```
D0485  GATCTGTGTTCTTCAGAAGTGTCAAAGTCATGAAAGACAAGGAAAAACTG   50
D0485  AGGAGCTGTCACAGGTGGGAGGGGATTAAAGAGATATAACAACTAAATTC  100
D0485  ACTGTGGTACAGGCATACTTTGAAGATATTGTGGGTTTGGTTCTAGACCA  150
D0485  CTGCAATAAAGCAAATTTCACAATAAAGTGAGTCATATAATATTTTAGGT  200
D0485  TTCCCAGTGCATGTAAAAGTCATAACCTATTAAGTGTGCAATAGCATGAT  250
D0485  ATTTAAAAAATCATATATATCATATTTACAAATACTGTATTGCTAAAAAT  300
D0485  GTTAATGATTATTTGAGCCTTTACTGATGGAAAGGCTTGCCTCAATGTTG  350
D0485  ATGGCTACTGATTGATCAGGGTGGTGATTGCTGAAGGTTGGAGTGGCTGT  400
D0485  GGCAATTTTTAACAAGACAACAGTGAAGTTTGCCTCATTGACTGATTCTT  450
D0485  CCTTTCATGAAAAATTTCTCTGTTGCATGTAATGCTGTTCGATAGCATTT  500
D0485  TACCTAAGTAGGCCTTTTTTTTTTTTTTTCCAAATTGGAGTCAATCCTC  550
D0485  TCAAACTCTGCTGCTGCTTTATCAACTAAGTTTATGGAACATTCCCAGTG  600
D0485  TGTTACTTACCTCAACTCTTAGCTTGTTGGGGACGGTAACCGGGACCCAG  650
kerat                        ---C--------------------*-

D0485  TG*TCTGCTCCTGTCACCTTCGCCTCCT*AATCCCTAGCCACTATGCGTG  698
kerat  --C------------G------------G------------*--------

D0485  AATGCGTCTCCATCCACGTTGGCCAGGCTGGTGTCCAGATTGGCAATGTC  748
kerat  -G---A-----------------------------------------C-

D0485  TGCTGGGAGCTCTACTGCCTGGAACATGGCATCCAGCCCGATGGCCAGAT  798
kerat  -------------------------C-----------------------

D0485  GCCAAGTGACAAGACCATTGGGGGAGGAGATGACTCCTTCAACACCTT**  846
kerat  ---------------------------------------------CT
```

```
D0485  *CAGTGAGACGGGTGCTGGCAAGCATGTGCCCCGGGCTGTGTTTGTAGAC  895
kerat  T-----------C-----------C----------------------

D0485  TTGGAACCCATGGTCATTGATGAAGTTTGCACTGGCACCTACCGCCAGCT  945
kerat  ----------CA--------------C--------------------

D0485  CTTCCACCCTGAGCAGCTCATCACAGGCAAGGAAGATGCTGCCAATAACT  995
kerat  ---------------------------------------------

D0485  ATGCCCGAGGGCACTACACCATTGGCAAGGAGATCATTGACCTTGTGTTG 1045
kerat  ---------------------------------------------

D0485  GACCGAATTCGCAAGCTGGCTAACCAGTGCACTGGTTTTCAGGGCTTCTT 1095
kerat  --------------------G----------CC--C------------

D0485  GGTTTTCCACAGCTTTGGTGGGGGAACTGGTTCTGGGTTCACCTCCCTGC 1145
kerat  ---------------------------------------------

D0485  TCATAGAACGTCTCTTAGTTGATTATGGCAAGAAGTCCAAGCTGGAGTTC 1195
kerat  ----G-----C--G-C-----------------A---------------

D0485  TCCAATTACCCAGCGCCCCAGGTTTCCACAGCTGTAGTTGAGCCCTACAA 1245
kerat  ----T------G--A-----------------------------------

D0485  CTCCATCCTCACCACCCACACCACCCTGGAGCACTCTGATTGTGCCTTCA 1295
kerat  ---------------------------------------------

D0485  TGGTAGACAATGAGGCCATCTGTGACATCTGTTGTAGAAACCTCAATATC 1345
kerat  --------------------A----------C-----------G-----

D0485  GAGCGCCCAACCTACACTAACCTTAACCACCTTATTAGCCAGATTGTGTC 1395
kerat  ----------------------------G---------------------

D0485  CTCTATCACTGCTTCCCTGAGATTTGATGGAGCCCTGAATGTTGACCTGA 1445
kerat  ---C-----------------------------------------

D0485  CAGAATTCCAGACCAACCTGGTACCCTACTCCCACATCCACTTCCCTCTG 1495
kerat  ----------------------C------C---G---------------

D0485  GCGACATATGCCCCTGTCATTTCTGCTGAGAAAGCCTACCATGAACAGCT 1545
kerat  --C-----------------C-----------------------------

D0485  TTCTGTAGCAGAGATCACCAATGCTTCCTTTGAGCCAGCCAACCAGATGG 1595
kerat  ------------C-------------G----------------------

D0485  TGAAATGTGACCCTCGCCATGGTAAATACATGGCTTGCTGCCTATTGTAC 1645
kerat  --------------G---------------------------G------

D0485  CATGGTGATGTGGTTCCCAAAGATGCCAATGCTGCCATTGCCACCATCAA 1695
kerat  -G------C----------------T-----------------------

D0485  GACCAAGCGTAGCATCCAGTTTGTAGATTGGTGCCCCACTGGCTTCAAGG 1745
kerat  A--------C-CG-----------G-----------------------

D0485  TTGGCATCAACTACCAGTCTCCCACTGTGGTGCCTGGTGGAGACCTGGCC 1795
kerat  ----------------C---------------------------------

D0485  ATGGTACAGA*******GTGCATGCTGAGCAACACCACAGCCATTGCTGA 1838
kerat  -A--------GAGCTGT---------------------------------
```

```
D0485  GGCCTGGGCTCGCCTGGACCACAAGTTTGACCTGATGTATGCCAAGCGTG 1888
kerat  --------------------------------------------------

D0485  CCTTTGTTCACTGGTACGTGGGTGAGGGGATGGAGGAAGGCGAGTTTTCA 1938
kerat  --------------------------------------------------

D0485  GAGGCCCGTGAAGACATGGCTGCCCTTGAGAAGGATTATGAGGAGGTTGG 1988
kerat  ----------T---------------------------------------

D0485  TGTGGATTCTGTTGAAGGAGAGGGTGAGGAAGAAAGAGAGGAATACTA** 2036
kerat  ----------------------------------G-------------AT

D0485  *ATCCATTCCTTTTGGCCCTGCAGCATGTCGTGCTCCCAGAATTTCAGCT 2085
kerat  T----------------------------A------------------

D0485  TCAGCTTAGCTGACAGACATTAAAGCTTTCTGGTTAGA***TTTTCATTT 2132
kerat  --------A--------TG------------------TTG------C--

D0485  GGTGATCATGTCTTTTCCATGTGTACCTGTAATGTTTTTCCATCATGTCT 2182
kerat  --------------------------------A------------A---

D0485  CAAAGTACAGTCATTAACATAAAAAAAAAAGTTTATGTTATATTTTAGAT 2232
kerat  -------A------------C-

D0485  CATGTGGTGGCATTTCAACAATGTTCACAGYATCTTCACTAAGAGTCGAT 2282

D0485  TGGGTCTCAAGAAMCCACTTTTTTCTCCCATCCAGAGGAAGAAGCTTCTC 2332

D0485  ATCCACTCAAGTYTCATTATGAGATTGSAGCAATTCAGTCACATCTTCAG 2382

D0485  GCTCCACTTCTAATTGTAGCTCACTTGCTATTTCCACATCTGCAGTTACT 2432

D0485  TCCTCTAATAATGTCTTGAATACTCAAAGTCATCAACTTCTTCAAACTCC 2482

D0485  TGTTATGGTTGACATTTTGACCTTCTCTCATAAATCACTAATGCTCTTAA 2532

D0485  TGACATCTTAGAATGGTTAATTATTTCCCAAAGTTTTTCAACTTGTTTAG 2582

D0485  TCAAGATCCATCAGAAGATTCACAATTTATATGGAAGCTATAGCCTTATG 2632

D0485  AAATGTACTGCTGAACAAATGAGACTTGAAAATTGAAATTTCTCTTTGAT 2682

D0485  C  2683
```

Figure 5.8 Translation of D0485 α-tubulin gene sequence. Amino acids are abbreviated using the IUB nomenclature codes. The open reading frame indicated codes for a protein of 374 amino acids.

```
D0485  GATCTGTGTTCTTCAGAAGTGTCAAAGTCATGAAAGACAAGGAAAAACTG   50
D0485  AGGAGCTGTCACAGGTGGGAGGGGATTAAAGAGATATAACAACTAAATTC  100
D0485  ACTGTGGTACAGGCATACTTTGAAGATATTGTGGGTTTGGTTCTAGACCA  150
D0485  CTGCAATAAAGCAAATTTCACAATAAAGTGAGTCATATAATATTTTAGGT  200
D0485  TTCCCAGTGCATGTAAAAGTCATAACCTATTAAGTGTGCAATAGCATGAT  250
D0485  ATTTAAAAAATCATATATATCATATTTACAAATACTGTATTGCTAAAAAT  300
D0485  GTTAATGATTATTTGAGCCTTTACTGATGGAAAGGCTTGCCTCAATGTTG  350
D0485  ATGGCTACTGATTGATCAGGGTGGTGATTGCTGAAGGTTGGAGTGGCTGT  400
D0485  GGCAATTTTTAACAAGACAACAGTGAAGTTTGCCTCATTGACTGATTCTT  450
D0485  CCTTTCATGAAAAATTTCTCTGTTGCATGTAATGCTGTTCGATAGCATTT  500
D0485  TACCTAAGTAGGCCTTTTTTTTTTTTTTTTCCAAATTGGAGTCAATCCTC  550
D0485  TCAAACTCTGCTGCTGCTTTATCAACTAAGTTTATGGAACATTCCCAGTG  600
D0485  TGTTACTTACCTCAACTCTTAGCTTGTTGGGGACGGTAACCGGGACCCAG  650
D0485  TGTCTGCTCCTGTCACCTTCGCCTCCTAATCCCTAGCCACTATGCGTGAA  700
                                                M  R  E
D0485  TGCGTCTCCATCCACGTTGGCCAGGCTGGTGTCCAGATTGGCAATGTCTG  750
       C  V  S  I  H  V  G  Q  A  G  V  Q  I  G  N  V  C
D0485  CTGGGAGCTCTACTGCCTGGAACATGGCATCCAGCCCGATGGCCAGATGC  800
        W  E  L  Y  C  L  E  H  G  I  Q  P  D  G  Q  M  P
D0485  CAAGTGACAAGACCATTGGGGGAGGAGATGACTCCTTCAACACCTTCAGT  850
         S  D  K  T  I  G  G  G  D  D  S  F  N  T  F  S
D0485  GAGACGGGTGCTGGCAAGCATGTGCCCCGGGCTGTGTTTGTAGACTTGGA  900
       E  T  G  A  G  K  H  V  P  R  A  V  F  V  D  L  E
D0485  ACCCATGGTCATTGATGAAGTTTGCACTGGCACCTACCGCCAGCTCTTCC  950
        P  M  V  I  D  E  V  C  T  G  T  Y  R  Q  L  F  H
D0485  ACCCTGAGCAGCTCATCACAGGCAAGGAAGATGCTGCCAATAACTATGCC 1000
         P  E  Q  L  I  T  G  K  E  D  A  A  N  N  Y  A
D0485  CGAGGGCACTACACCATTGGCAAGGAGATCATTGACCTTGTGTTGGACCG 1050
       R  G  H  Y  T  I  G  K  E  I  I  D  L  V  L  D  R
```

```
D0485  AATTCGCAAGCTGGCTAACCAGTGCACTGGTTTTCAGGGCTTCTTGGTTT 1100
         I  R  K  L  A  N  Q  C  T  G  F  Q  G  F  L  V  F

D0485  TCCACAGCTTTGGTGGGGGAACTGGTTCTGGGTTCACCTCCCTGCTCATA 1150
         H  S  F  G  G  G  T  G  S  G  F  T  S  L  L  I

D0485  GAACGTCTCTTAGTTGATTATGGCAAGAAGTCCAAGCTGGAGTTCTCCAA 1200
         E  R  L  L  V  D  Y  G  K  K  S  K  L  E  F  S  N

D0485  TTACCCAGCGCCCCAGGTTTCCACAGCTGTAGTTGAGCCCTACAACTCCA 1250
         Y  P  A  P  Q  V  S  T  A  V  V  E  P  Y  N  S  I

D0485  TCCTCACCACCCACACCACCCTGGAGCACTCTGATTGTGCCTTCATGGTA 1300
         L  T  T  H  T  T  L  E  H  S  D  C  A  F  M  V

D0485  GACAATGAGGCCATCTGTGACATCTGTTGTAGAAACCTCAATATCGAGCG 1350
         D  N  E  A  I  C  D  I  C  C  R  N  L  N  I  E  R

D0485  CCCAACCTACACTAACCTTAACCACCTTATTAGCCAGATTGTGTCCTCTA 1400
         P  T  Y  T  N  L  N  H  L  I  S  Q  I  V  S  S  I

D0485  TCACTGCTTCCCTGAGATTTGATGGAGCCCTGAATGTTGACCTGACAGAA 1450
         T  A  S  L  R  F  D  G  A  L  N  V  D  L  T  E

D0485  TTCCAGACCAACCTGGTACCCTACTCCCACATCCACTTCCCTCTGGCGAC 1500
         F  Q  T  N  L  V  P  Y  S  H  I  H  F  P  L  A  T

D0485  ATATGCCCCTGTCATTTCTGCTGAGAAAGCCTACCATGAACAGCTTTCTG 1550
         Y  A  P  V  I  S  A  E  K  A  Y  H  E  Q  L  S  V

D0485  TAGCAGAGATCACCAATGCTTCCTTTGAGCCAGCCAACCAGATGGTGAAA 1600
         A  E  I  T  N  A  S  F  E  P  A  N  Q  M  V  K

D0485  TGTGACCCTCGCCATGGTAAATACATGGCTTGCTGCCTATTGTACCATGG 1650
         C  D  P  R  H  G  K  Y  M  A  C  C  L  L  Y  H  G

D0485  TGATGTGGTTCCCAAAGATGCCAATGCTGCCATTGCCACCATCAAGACCA 1700
         D  V  V  P  K  D  A  N  A  A  I  A  T  I  K  T  K

D0485  AGCGTAGCATCCAGTTTGTAGATTGGTGCCCCACTGGCTTCAAGGTTGGC 1750
         R  S  I  Q  F  V  D  W  C  P  T  G  F  K  V  G

D0485  ATCAACTACCAGTCTCCCACTGTGGTGCCTGGTGGAGACCTGGCCATGGT 1800
         I  N  Y  Q  S  P  T  V  V  P  G  G  D  L  A  M  V

D0485  ACAGAGTGCATGCTGAGCAACACCACAGCCATTGCTGAGGCCTGGGCTCG 1850
         Q  S  A  C  *

D0485  CCTGGACCACAAGTTTGACCTGATGTATGCCAAGCGTGCCTTTGTTCACT 1900

D0485  GGTACGTGGGTGAGGGGATGGAGGAAGGCGAGTTTTCAGAGGCCCGTGAA 1950

D0485  GACATGGCTGCCCTTGAGAAGGATTATGAGGAGGTTGGTGTGGATTCTGT 2000

D0485  TGAAGGAGAGGGTGAGGAAGAAAGAGAGGAATACTAATCCATTCCTTTTG 2050

D0485  GCCCTGCAGCATGTCGTGCTCCCAGAATTTCAGCTTCAGCTTAGCTGACA 2100

D0485  GACATTAAAGCTTTCTGGTTAGATTTTCATTTGGTGATCATGTCTTTTCC 2150
```

D0485 ATGTGTACCTGTAATGTTTTTCCATCATGTCTCAAAGTACAGTCATTAAC 2200

D0485 ATAAAAAAAAAAGTTTATGTTATATTTTAGATCATGTGGTGGCATTTCAA 2250

D0485 CAATGTTCACAGYATCTTCACTAAGAGTCGATTGGGTCTCAAGAAMCCAC 2300

D485 TTTTTTCTCCCATCCAGAGGAAGAAGCTTCTCATCCACTCAAGTYTCATT 2350

D0485 ATGAGATTGSAGCAATTCAGTCACATCTTCAGGCTCCACTTCTAATTGTA 2400

D0485 GCTCACTTGCTATTTCCACATCTGCAGTTACTTCCTCTAATAATGTCTTG 2450

D0485 AATACTCAAAGTCATCAACTTCTTCAAACTCCTGTTATGGTTGACATTTT 2500

D0485 GACCTTCTCTCATAAATCACTAATGCTCTTAATGACATCTTAGAATGGTT 2550

D0485 AATTATTTCCCAAAGTTTTTCAACTTGTTTAGTCAAGATCCATCAGAAGA 2600

D0485 TTCACAATTTATATGGAAGCTATAGCCTTATGAAATGTACTGCTGAACAA 2650

D0485 ATGAGACTTGAAAATTGAAATTTCTCTTTGATC 2683

Figure 5.9 Comparison of the peptide sequences of the D0485 α-tubulin ('D0485', upper line) and the keratinocyte α-tubulin ('kerat', lower line). The seven base pair deletion in the D0485 α-tubulin, which occurs within the codon for amino acid 372 results in a premature termination codon a further three amino acids downstream. The full length keratinocyte peptide is 451 amino acids long. In addition, a three base pair deletion in the D0485 α-tubulin sequence removes a phenylalanine amino acid at position 52. Homologous but not identical amino acids are represented by ':' . Termination codons are represented by '*'.

```
D0485  MRECVSIHVGQAGVQIGNVCWELYCLEHGIQPDGQMPSDKTIGGGDDSFN   50
       ||||:|||||||||||||:||||||||||||||||||||||||||||||||
kerat  MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFN   50

D0485  T-FSETGAGKHVPRAVFVDLEPMVIDEVCTGTYRQLFHPEQLITGKEDAA   99
       | |||||||||||||||||||| ||||| ||||||||||||||||||||||
kerat  TFFSETGAGKHVPRAVFVDLEPTVIDEVRTGTYRQLFHPEQLITGKEDAA  100

D0485  NNYARGHYTIGKEIIDLVLDRIRKLANQCTGFQGFLVFHSFGGGTGSGFT  149
       ||||||||||||||||||||||||||||||:||| :||||||||||||||||
kerat  NNYARGHYTIGKEIIDLVLDRIRKLADQCTRLQGFLVFHSFGGGTGSGFT  150

D0485  SLLIERLLVDYGKKSKLEFSNYPAPQVSTAVVEPYNSILTTHTTLEHSDC  199
       |||:||| |||||||||||| ||||||||||||||||||||||||||||||
kerat  SLLMERLSVDYGKKSKLEFSIYPAPQVSTAVVEPYNSILTTHTTLEHSDC  200

D0485  AFMVDNEAICDICCRNLNIERPTYTNLNHLISQIVSSITASLRFDGALNV  249
       |||||||||:||| |||:||||||||||:||||||||||||||||||||||
kerat  AFMVDNEAIYDICRRNLDIERPTYTNLNRLISQIVSSITASLRFDGALNV  250

D0485  DLTEFQTNLVPYSHIHFPLATYAPVISAEKAYHEQLSVAEITNASFEPAN  299
       ||||||||||||::||||||||||||||||||||||:||||:||||||||
kerat  DLTEFQTNLVPYPRIHFPLATYAPVISAEKAYHEQLSVADITNACFEPAN  300

D0485  QMVKCDPRHGKYMACCLLYHGDVVPKDANAAIATIKTKRSIQFVDWCPTG  349
       ||||||| ||||||||||:|||||||:||||||||||:||||||||||||
kerat  QMVKCDPGHGKYMACCLLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTG  350

D0485  FKVGINYQSPTVVPGGDLAMVQSAC*----------------------- 374
       ||||||||:||||||||||:||:|
kerat  FKVGINYQPPTVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYA 400

D0485  ---------------------------------------------------

kerat  KRAFVHWYVGEGMEEGEFSEAREDMAALEKDYEEVGVDSVEGEGEEEGEE 450

D0485  --

kerat  Y*   451
```

5.3 Expression of the D0485 Alpha Tubulin Gene

The fact that the D0485 α-tubulin gene has no introns and that alterations in the coding sequence would lead to the production of a truncated peptide favours the possibility that this sequence is a pseudogene, resulting from the integration into the genome of a reverse transcribed keratinocyte α-tubulin mRNA species. On the other hand, examples of intronless expressed genes have been reported in the literature. The dopamine D1 receptor is intronless (Sunahara *et al*, 1990) as are histone genes (Stein, 1984) and type X collagen (Ninomiya *et al*, 1986) among others. Furthermore, the depressed mutation rate within the coding sequence relative to the UTR regions may suggest some functional evolutionary pressure for selective retention of the coding sequence.

5.3.1 Development of an allele specific assay for the D0485 α-tubulin gene

Due to the similarity between the D0485 α-tubulin gene and other α-tubulin genes, it would be impossible to investigate the expression of this particular gene by hybridisation of the CSC tubulin clone probes to Northern blots. Hybridisation of an allele-specific oligonucleotide probe could resolve this problem, yet determination of extremely specific hybridisation and washing conditions would be required. Even so, a negative result would be difficult to interpret and a positive result difficult to verify. Hence an allele-specific RT-PCR assay was devised.

This assay employed two oligonucleotides, 496 and 497 (see table 2.2), which were specific to the D0485 α-tubulin gene. The final three 3' bases of oligonucleotide 496 spanned the deletion of CTT from the keratinocyte α-tubulin sequence, and the final three 3' bases of 497 spanned the seven base pair deletion of from the keratinocyte α-tubulin sequence. When

employed together as primers in a PCR reaction, the target product was 1013bp, derived specifically from the D0485 α-tubulin sequence.

In order to test the assay, the PCR reaction was firstly performed on D0485 YAC DNA and total human genomic DNA. 300ng of each primer was used to amplify 50ng of YAC DNA and 200ng of total human genomic DNA. The PCR programme used was:
Hot start at 90°C
Denaturation: 94°C for 30 secs in the first cycle and 15 secs thereafter
Annealing: Touch down from 66°C to 56°C then 56°C, for 30 secs
Extension: 72°C for 1 min (cycles 1-10), 2 mins (11-20) and 3 mins (21-30).

A strong band of the expected size was obtained for both YAC DNA and total human genomic DNA. In order to verify that this product was indeed derived specifically from the D0485 α-tubulin gene, the two ends of the product were sequenced by radioactive sequencing of the PCR product. Oligonucleotides 714 and 715 were designed for this purpose (see table 2.2) - both lie within the amplified sequence, orientated to permit sequencing across the very end of the PCR product, where the two deletions are situated. Sequencing of these two ends confirmed that the product was derived from the D0485 α-tubulin sequence.

5.3.2 Development of positive control RT-PCR assays

Total RNA was extracted from a variety of human tissues. The tissues used were adult brain (two different samples), colon, liver, prostate and testis, and foetal brain (gestational ages 10 and 16.5 weeks), spleen, liver, limb, heart and kidney. The RNA from these samples was tested by a positive control PCR to amplify exons 3 to 8 of the ubiquitously expressed glutaraldehyde phosphate dehydrogenase (GAPDH) gene. 300ng of the forward and reverse GAPDH primers (see table 2.2) were used to amplify 600ng of DNase-treated RNA from each tissue. This reaction employed the same

PCR programme as in the allele-specific α-tubulin assay as above, so that a GAPDH control could be included with every α-tubulin RT-PCR reaction carried out. A minus reverse transcriptase enzyme control was also included, to ensure that any resulting product was derived from RNA and not DNA contamination.

A band of the expected size (442 bp) was produced from all tissues. This PCR reaction could therefore be used as a positive control, both for the integrity of the RNA and for the RT-PCR reaction itself.

A second positive control, to assay for expression of other α-tubulin genes, was also carried out on 600ng of both gestational ages of foetal brain DNase-treated RNA. 300ng of each primer was used.
The PCR programme was:
Denaturation: 94°C for 30 secs in the first cycle and 15 secs thereafter
Annealing: Touch down from 65°C to 55°C then 55°C, for 30 secs
Extension: 72°C for 30 secs (cycles 1-10), 45 secs (11-20) and 1 min (21-30)

Two primer combinations were used, 152/312 and 156/158 (see table 2.2). Primers 156 and 158 are both situated within the α-tubulin coding sequence, which is conserved between subtypes, and therefore the target 531bp 156/158 PCR product was not subtype-specific. Primer 152 however lies at the 5' end of the gene, slightly upstream of the beginning of the keratinocyte α-tubulin sequence. If the target 690bp 152/312 PCR product were generated, then it would be specific to subtypes with a 5' UTR very similar to that of keratinocyte α-tubulin. For both PCR reactions, a product of the expected size was produced from foetal brain RNA. The intensity of the band was approximately equal to that produced by the GAPDH control RT-PCR, which indicates that the D0485 α-tubulin specific RT-PCR should be sufficiently sensitive to detect expression if this gene is expressed at a similar level to the keratinocyte α-tubulin gene.

5.3.3 Allele-specific RT-PCR of the D0485 α-tubulin gene

RT-PCR from the aforementioned human tissues was then carried out using the allele-specific α-tubulin primers 496 and 497 to amplify 600ng of DNase-treated RNA. The PCR programme used was that employed in the allele-specific α-tubulin assay in section 5.3.1. An additional PCR performed on total human genomic DNA provided a control for the primers and other components of the PCR reaction mix. A set of minus reverse transcriptase PCRs controlled for the presence of DNA in the RNA samples. Finally a PCR using the GAPDH primers (plus and minus reverse transcriptase) acted as a control for the quality of the RNA and the efficiency of the first strand cDNA synthesis.

Figure 5.10 shows the results of this RT-PCR reaction on four of the tissues. The control PCRs from DNA and for GAPDH produced strong bands, yet there was no product for D0485 α-tubulin from RNA. The same results were obtained for the other eight tissues tested. It must be concluded that the D0485 α-tubulin gene is not expressed in any of the tissues tested, and is therefore most likely to be a pseudogene.

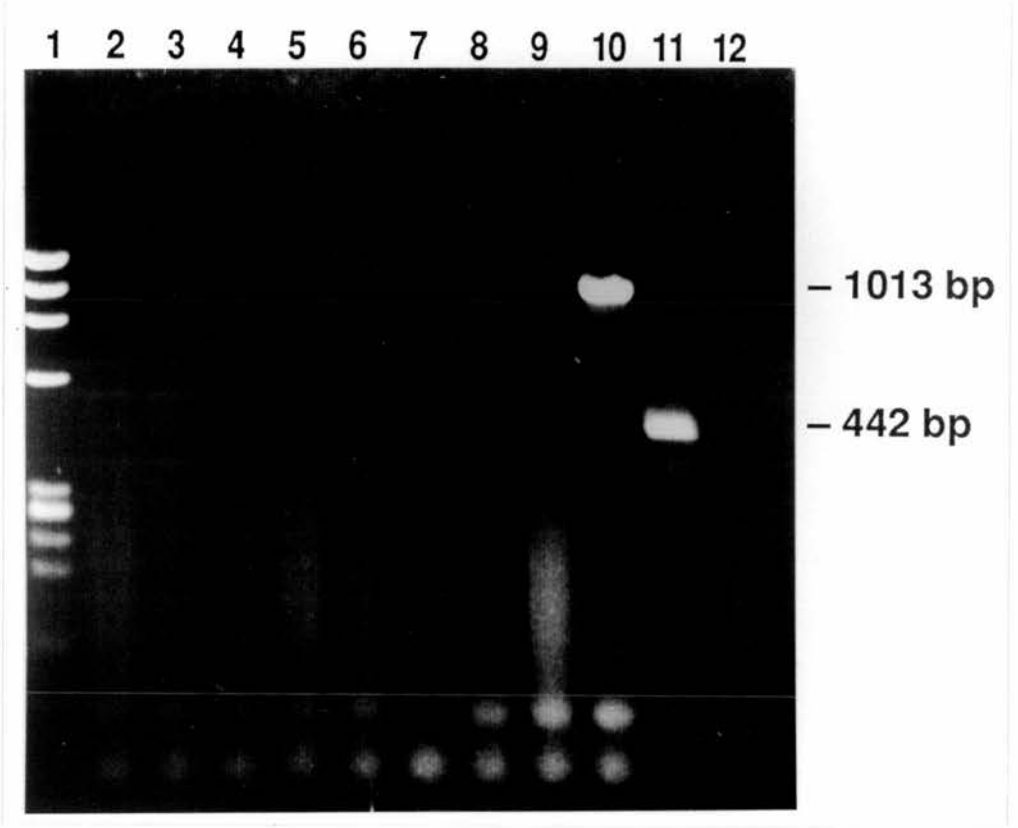


Figure 5.10 Allele specific RT-PCR of a fragment of the D0485 α-tubulin gene. The size marker used is 250ng of ϕx *Hae III* (track 1). Tracks 2 to 12 contain 10μl of PCR product resolved on a 1.5% agarose gel. Tracks 2 to 9 resulted from RT-PCR on DNase-treated RNA from foetal brain (tracks 2 and 6), foetal spleen (tracks 3 and 7), foetal heart (tracks 4 and 8) and adult liver (tracks 5 and 9) with primers 496 and 497. Tracks 6 to 9 are a minus reverse transcriptase negative control. The band at 1013 bp in track 10 was the result of amplification of total human genomic DNA with the same primers 496 and 497. The band at 442 bp in track 11 resulted from RT-PCR of a portion of the GAPDH gene from foetal brain DNase-treated RNA. Track 12 contains a minus reverse transcriptase control for the GAPDH RT-PCR. The two positive controls in tracks 10 and 11 imply that the absense of a band in tracks 2 to 5 is genuinely due to non-expression of the D0485 α-tubulin gene in these tissues. The same result was obtained from the other tissues tested.

5.4 Discussion

The lack of introns in the D0485 α-tubulin gene and the lack of expression in a variety of normal foetal and adult tissues does suggest that this sequence is a processed pseudogene, resulting from the re-insertion into the genome of an α-tubulin mRNA species. The close similarity between the sequence of the D0485 gene and the published keratinocyte α-tubulin gene suggests that the former is derived from the latter, and that the event occurred fairly recently in evolutionary terms.

Contrary to the idea that the D0485 α-tubulin sequence is a pseudogene is the finding that there are more mutations relative to the keratinocyte sequence in the 5' and 3' UTRs than in the coding sequence; regions which are subject to less functional constraint. If a protein were translated from the D0485 α-tubulin gene, it would be truncated to 80% of its normal length (374 amino acids rather than 451 in keratinocyte α-tubulin). The 77 deleted amino acids are at the C-terminus. The extreme C-terminal end of the α-tubulin protein has been implicated in the correct formation of microtubules (Arévalo *et al*, 1990), binding of ATP (Jayaram and Haley, 1994) the binding of microtubule associated proteins (Paschal *et al*, 1989) and some post-translational modifications (Boucher *et al*, 1994). It must be assumed that these aspects of tubulin function would be disturbed by the C-terminal truncation. However, formation of α-β tubulin dimers (Arévalo *et al*, 1990), other post-translational modifications (Boucher *et al*, 1994), G-protein and GTP binding (Jayaram and Haley, 1994) and the binding of the ATPase translocators kinesin and dynein (Rodionov *et al*, 1990) would be unaffected.

Although expression of the D0485 α-tubulin gene was not detected in the wide range of tissues tested, it is possible that this particular gene is expressed, but in a highly specific manner, perhaps even developmentally regulated. Tubulin heterogeneity is higher in the brain and neurons than in other cell types, and thus implies the existence of tubulin isoforms that are

expressed predominantly or specifically in neurons (Audebert *et al*, 1994). If the D0485 α-tubulin gene were involved in the aetiology of schizophrenia, a brain/neuron specific expression pattern may be expected.

It is possible that the D0485 α-tubulin gene may be aberrantly expressed in individuals bearing the t(1;11) translocation. Although it is approximately 300kb from the translocation breakpoint, no gene has been identified from chromosome 11 which maps closer, and the pattern of rare cutter restriction enzyme sites, indicative of CpG islands, suggests that one will not be found. The translocation could bring new control elements into the vicinity of the tubulin gene, which may exert an effect on its expression via a position effect. Other cases have been described in which a cytogenetic abnormality at a distance from a known gene is associated with the phenotype associated with mutations in that gene. For instance, two pedigrees have been described in which chromosomal rearrangements co-segregate with aniridia (Fantes *et al*, 1995). In both cases the chromosomal breakpoint is at least 85kb from the Pax 6 gene (which is known to be mutated in other aniridia patients) while the gene itself is apparently free of mutations. A dominant negative effect of the truncated form is also possible, if the der(11) chromosome but not the normal chromosome 11 version is expressed and active in particular cells at a critical stage of development.

Allele-specific RT-PCR of the D0485 α-tubulin gene on brain tissue derived from individuals carrying the translocation or somatic cell hybrid cells containing the der(1) or der(11) chromosomes only could assess whether the translocation event causes the aberrant expression of this gene. Complicating factors such as specificity of expression to certain brain regions or even a certain developmental stage may however preclude this. A more direct approach would be to transfer the der (11) translocation chromosome into suitable cells (for instance rat glioma) in culture, and then to assay for the transcription of the D0485 α-tubulin gene *in vitro*.

Although it has not yet been specifically proposed, a rôle for a tubulin gene in the aetiology of schizophrenia can be envisaged. Abnormalities in cytoarchitecture, of particularly the hippocampus, have long been associated with schizophrenia, and thus it is of interest to note that tubulin plays a major rôle in the architecture of the cell, via microtubules. Microtubules themselves have also been implicated in schizophrenia (Kerwin, 1993); an abnormality in microtubule associated proteins (MAPs) may cause a disruption in microtubule assembly and a consequent developmental anomaly in the temporal lobe. Abnormal expression of MAP2 and MAP5 has been detected in the hippocampus of some schizophrenics (Arnold *et al*, 1991). Another MAP, tau, has been implicated in Alzheimer's disease. Tau, which is thought to stabilise microtubules in axons, forms the major component of the neurofibrillary tangles found in the brains of patients with Alzheimer's disease (Mandelkow and Mandelkow, 1995).

In general, microtubules are involved in numerous aspects of brain development and function, including the maintenance and growth of axons, intracellular transport along axons and the determination of cellular morphology (Cambray-Deakin and Burgoyne, 1987). All of the above, especially the former, have been described as abnormal in subsets of schizophrenics. Neuronal tubulin may play a fundamental rôle in these processes, perhaps by the selective stabilisation of certain microtubules (Mandelkow and Mandelkow, 1995).

A link can also be forged between tubulin and the functioning of neurotransmitter systems. Components of the cytoskeleton are important in the localisation of membrane components to post-synaptic sites, and $GABA_A$ receptors probably bind to microtubules directly via tubulin (Item and Sieghart, 1994). Post-translational modification of α- and β-tubulin in nerve cells is very widespread, particularly addition of a short polyglutamate chain

at the carboxy end of the protein (Audebert *et al*, 1994). Although the functional significance of this is not known, increasing length of this chain is a sign of neuronal differentiation and probably regulates the binding of tau (Boucher *et al*, 1994). An abnormally low level of glutamate, as has been proposed to occur in schizophrenia, may prohibit the formation of this glutamate chain and therefore interfere with normal neuronal development.

It has also been proposed that dimeric tubulin plays a rôle in signal transduction, through its association with the GTP-binding (G) proteins G_s and G_{ii} (Roychowdhury and Rasenick, 1994). The α-subunits of some G-proteins form complexes with dimers of synaptic membrane tubulin molecules, and transfer a nucleotide from the GTP bound to tubulin. This may create an interdependence between the shape of the synapse and the response to neurotransmitters (Roychowdhury and Rasenick, 1994). The phosphorylation of cytoskeletal proteins may also be important in the non-receptor-dependent regulation of adenyl cyclase activity by calmodulin, since calmodulin has been reported to stimulate adenyl cyclase exclusively in neuronal systems (Rasenick *et al*, 1990). The GTP and G-protein binding capabilities of tubulin, which would be most important in modulation of neurotransmitter function, are also those which would be unaffected by a C-terminal truncation, as occurs in the D0485 α-tubulin gene.

CHAPTER 6
ANALYSIS OF THE NOVEL cDNA FRAGMENTS

From the analysis of the CSC product libraries, five families and two single products of novel cDNA fragments had been identified. Since the sequence of these products gave no significant database matches, nothing could be further deduced about the total number of genes from which they were derived or any aspect of their function. It was firstly necessary to prove the authenticity of the products by hybridisation back to YAC D0485 and locate their mapping positions within the YAC. Studies could then be undertaken to assess the copy number of these fragments in the human genome, to group the fragments into different genes and to begin to assess the expression pattern of these genes.

6.1 Hybridisation of cDNA Fragments to YAC D0485

One cDNA fragment from each family (that identified first) plus the two single fragments (in total, nonran4, 2-10b, 5-3c, 4-7d, 1-11d, 4-1f and 3-3a) were used as hybridisation probes onto filters containing restriction fragments from YAC D0485, resolved by pulsed field gel electrophoresis (made by John Maule). These filters contained *Sal I*, *Sfi I* and *Not I* digests, plus double digests thereof. All of the products gave positive signals on the YAC, confirming that they were indeed cDNA fragments derived from it. Figure 6.1 shows the results obtained with probes 2-10b and 4-1f. Table 6.1 shows the sizes of the restriction fragments to which each of the probes hybridised.

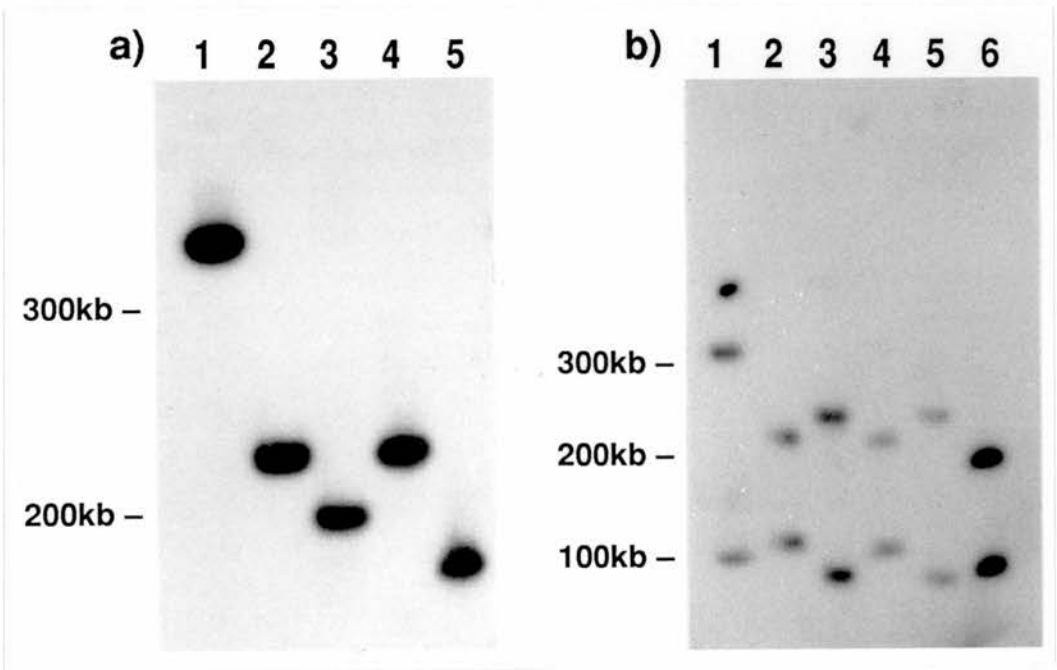
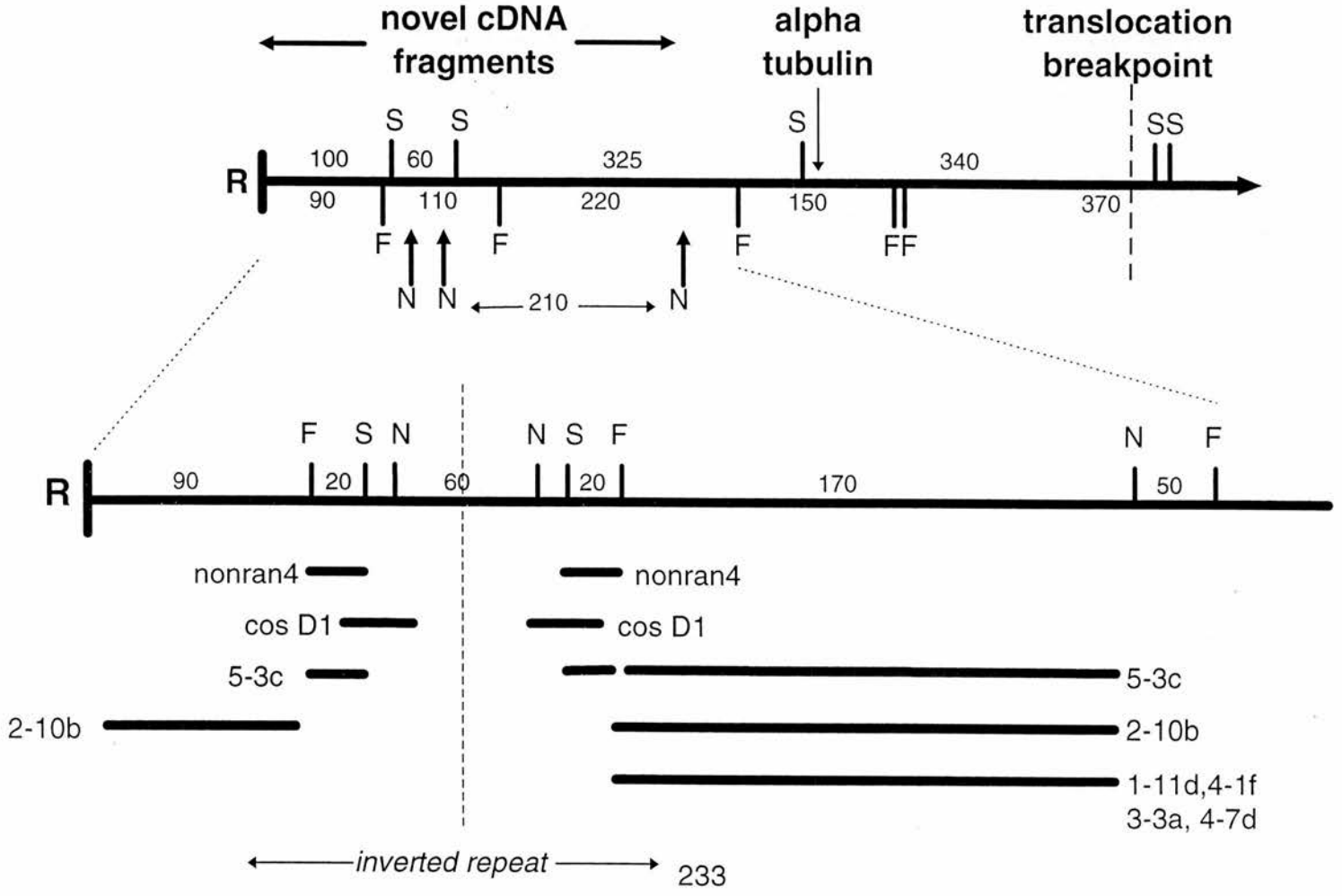


Figure 6.1 Hybridisation of probes 4-1f (left) and 2-10b (right) to restriction digested YAC D0485. For 4-1f tracks 1 to 5 correspond to *Sal I*, *Sfi I*, *Sal I* & *Not I*, *Sal I* & *Sfi I* and *Sfi I* & *Not I* digests respectively. For 2-10b tracks 1 to 6 correspond to *Sal I*, *Not I*, *Sfi I*, *Sal I* & *Not I*, *Sal I* & *Sfi I* and *Not I* & *Sfi I* digests respectively. The sizes of the hybridising bands enabled placement of the novel cDNA fragment probes on the long-range restriction map of the YAC.

Table 6.1 ND= not determined						
	Restriction fragment sizes to which the probes hybridised (kb)					
Probe	**Sal I**	**Sfi I**	**Not I**	**Sal I / Sfi I**	**Sfi I / Not I**	**Sal I / Not I**
nonran4	325, 100	120	210, 110	20	20	210, 100
2-10b	325, 100	230, 90	210, 110	230, 90	90, 170	100, 210
5-3c	325, 100	230, 100	ND	20, 230	30, 170	100, 210
4-7d	325	230	210	230	170	210
1-11d	325	230	210	230	170	210
4-1f	325	230	ND	230	170	210
3-3a	325	230	210	230	170	210

The results from hybridisation of these fragments onto PFGE filters enabled location of their position on the long range restriction map of the YAC. All the fragments were derived from the centromeric 'right hand' end of the YAC, proximal to the translocation breakpoint (Fig. 6.2). This is the same region of the YAC as that which contains the *Not I* restriction sites, which are indicative of CpG islands and therefore genes. Finer scale mapping was required however before any definite association of the cDNA fragments with the *Not I* sites could be asserted. Probes nonran4, 2-10b and 5-3c hybridised to two different restriction fragments in each track. Combining this information with data generated from the mapping of two fragments of the cosmid D1 (by John Brown; data not shown) suggested that the YAC D0485 contains an inverted repeat structure in this region. It is not equivocally proven whether this structure represents an artefact in the YAC, or whether it is present in normal genomic DNA. (Evidence to suggest the latter will be presented in section 6.3)

Figure 6.2 *Sal I* (S), *Sfi I* (F) and *Not I* (N) pulsed field restriction map of the right hand end of YAC D0485 showing the positions of the novel cDNA fragments in relation to the alpha tubulin gene and the translocation breakpoint. The enlarged portion shows the maximum region in which the individual fragments may lie, except cosmid D1 for which the exact position is shown. 4-7d, 1-11d, 4-1f and 3-3a lie within the same *Sfi I* & *Not I* restriction fragment and have therefore been depicted as a single bar. Nonran4 and 2-10b map to two places in the YAC, on either side of a predicted inverted repeat structure. 5-3c maps to three places. Numbers represent the distance between the restriction sites, in kb.



6.2 Hybridisation of Fragments to cDNA

As a first step towards proving the genic nature of these novel fragments, one product from each Family (except Family 5) was used as a hybridisation probe onto filters containing human foetal brain and mouse cDNA. The results are shown in Figure 6.3. All the probes hybridised to the cDNA except nonran4 (Family 1). This may imply that this product is genuinely not represented in these cDNA populations, or it may indicate merely a very low expression level that resulted in a signal too weak to see under these conditions. A very low expression level may explain why this Family is comprised of clones derived from the more stringent EL-CSC only. The probe 2-10b (Family 2) gave a strong signal on human cDNA but a much weaker signal on mouse cDNA. The other probes, 5-3c, 4-7d, 4-1f and 3-3a (Families 3, 4 and Products 6 and 7 respectively) all produced a similar result, allowing them to be tentatively ascribed to the same gene. They all produced strong signals on cDNA, with approximately equal intensity from human and mouse.

The signal from the total human genomic track included on the filters permits some speculation regarding the expression level of these products relative to each other. For instance, for probe 2-10b the cDNA signal is very strong, and the signal in the genomic DNA is also clear. For probe 5-3c however, the cDNA signal is strong, but the genomic DNA track is blank. This may imply that in fact 5-3c is more highly expressed than 2-10b, since a clear cDNA signal is visible before the genomic signal becomes apparent.

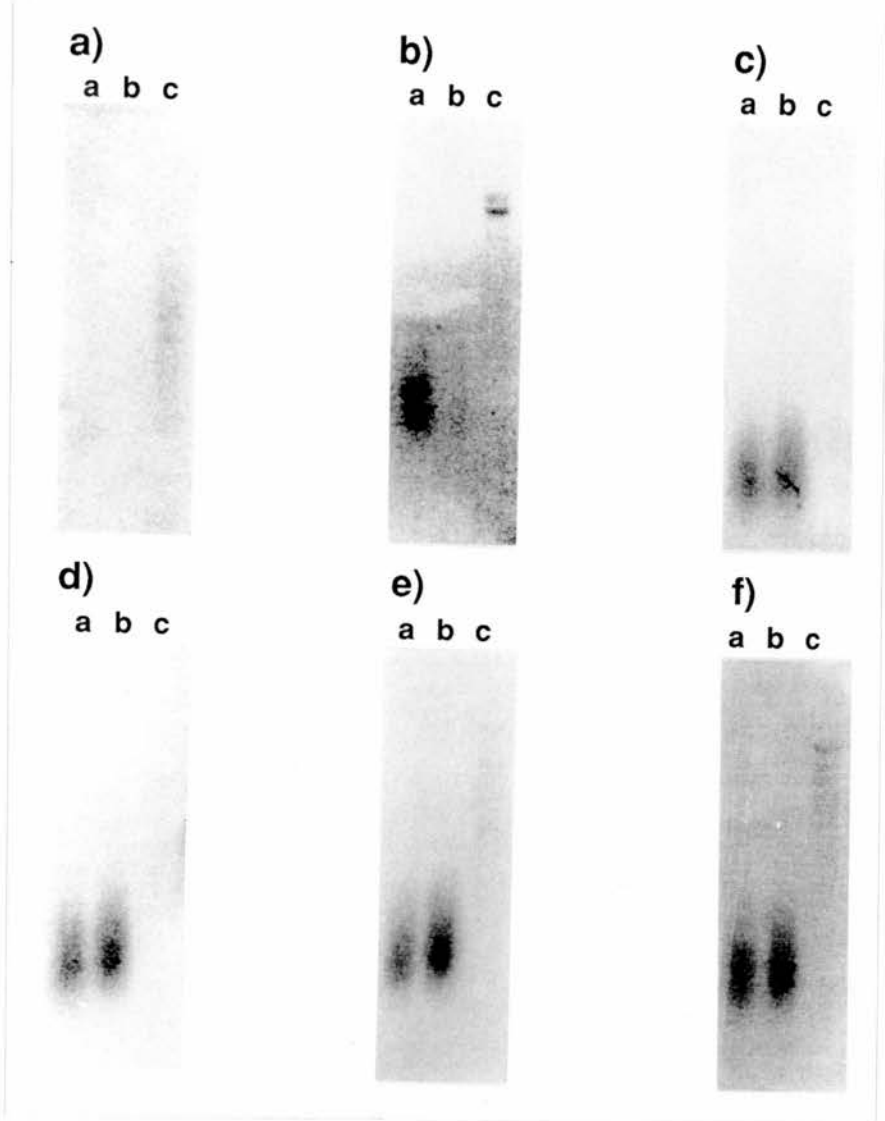


Figure 6.3 Hybridisation of six of the CSC novel fragment probes to (a) catch-linkered & amplified human foetal brain cDNA, (b) catch-linkered & amplified mouse adult brain cDNA, (c) total human genomic DNA digested with *EcoRI*. a = nonran4, b = 2-10b, c = 5-3c, d = 4-7d, e = 4-1f and f = 3-3a. Probe nonran4 gave no signal on cDNA, possibly due to a prohibitively low expression level. The other probes resulted in a smear of size ~200-700bp on cDNA. A smear of hybridisation signal rather than a band is a result of incomplete digestion of the cDNA, chimaeric products and non-specific priming in the PCR reaction. The relevance of the total human genomic track (c) has been discussed in the text.

6.3 Copy Number of the cDNA Fragments

In order to investigate the nature of the cDNA fragments with respect to their copy number in genomic DNA, the fragments nonran4, 2-10b, 5-3c, 4-7d, 1-11d, 4-1f and 3-3a were used as hybridisation probes onto filters containing total human genomic DNA digested with *EcoRI*, run parallel to total yeast plus YAC D0485, also digested with *EcoRI*. The amounts of human genomic DNA and yeast plus YAC DNA (5μg and 200ng respectively) were adjusted to give signals of comparable strength. Fig. 6.4 shows the results of these hybridisations.

Probe nonran4 hybridises to three (and perhaps four) bands in total human genomic DNA all between 9.5kb and 23kb. The slight background smear indicates that this probe contains a small region of a high copy repeat sequence. In the YAC track there was a single band, which was probably the same size as the largest genomic band, even though it appeared slightly larger due to the lower signal intensity of the genomic bands. The single YAC band could indicate a single copy of this sequence in the YAC; however the inverted repeat structure revealed by pulsed field analysis had previously demonstrated at least two copies. *EcoRI* fragments are a great deal smaller (average fragment size 3.2kb; Drmanac *et al*, 1986) than those which can be resolved by PFGE. Since the nonran4 sequence occurs within only one *EcoRI* fragment, there are likely to be only two copies within the YAC, unless the sequence occurs as an array of tandem repeats. Outwith the YAC D0485, nonran4 occurs at two (possibly three) other sites within the genome. The signal intensity of each genomic band is approximately equal, which indicates the same number of copies of nonran4 at each location.

Probe 2-10b hybridises to six bands in total human genomic DNA, all of relatively equal signal strength (suggesting equal number of copies of 2-10b, probably one at each locus). In YAC D0485, 2-10b hybridises to two bands,

of the same size as the smallest two genomic bands (approximately 5kb and 6.5kb).

Probe 5-3c hybridises to three bands of equal intensity in total human genomic DNA, and to three bands of the same sizes (all between 6 and 9.5kb) in the YAC DNA. This indicates that there are at least three copies of the 5-3c sequence within the YAC, and that the sequence does not occur at any other sites in the genome.

Probe 4-7d hybridises to two bands in total human genomic DNA, one at approximately 5kb and the other at approximately 15kb. The larger band is of a far greater intensity than the smaller, which suggests that this locus contains multiple copies of the 4-7d sequence. In the YAC however there is a single band, which corresponds in size to the smaller band in genomic DNA. This is consistent with a single copy of 4-7d on the YAC.

Probe 1-11d hybridises to four bands in total human genomic DNA, the upper two of which appear single copy and the lower two of which contain many copies of 1-11d. In the YAC, 1-11d hybridises to a single band, of the same size as the second largest band in genomic DNA (approximately 6kb). This suggests a unique copy of the sequence on the YAC, plus another single copy locus and two multi-copy loci elsewhere in the genome.

Probe 4-1f contains part of a high copy repeat sequence (this was revealed during the database searches) and as such produces an intense smear on human genomic DNA. The smear masks bands which could also be present, although there appears to be an intense (therefore multi-copy) band at approximately 7kb and a single copy band at 1.5kb. In the YAC, 4-1f hybridises to a single band of approximately 1.5kb.

Probe 3-3a gives exactly the same hybridisation pattern as probe 4-7d on total human genomic DNA and YAC D0485, which suggests that these two fragments lie in close proximity to each other.

The results from hybridisation of the cDNA fragment probes to total genomic DNA suggest that all the fragments are present in more than one copy in the genome. 5-3c is the most localised product, with three copies in the YAC, and no others elsewhere in the genome. For the fragments producing intense multi-copy bands on genomic DNA, it is impossible to tell without further analysis how many copies there are and whereabouts in the genome they are located. The multi-copy nature of the fragments suggests that the genes from which they are derived are members of gene families or that there are pseudogenes derived from them. Alternatively, the fragments could be derived from low copy repeat (MER) elements, which can be sufficiently sequence specific that they are mistakenly interpreted as a gene family. Further proof of the genic nature of the fragments (for instance demonstration of splicing and expression data) would be required to resolve this matter.

For all the cDNA fragments, there is a hybridising band in genomic DNA of the same size as every band present in the YAC. This implies that all the *EcoRI* fragments of the YAC in which the cDNA fragments are situated are present in the same form in genomic DNA. This is evidence to suggest that the inverted repeat structure in YAC D0485 is also present in uncloned genomic DNA.

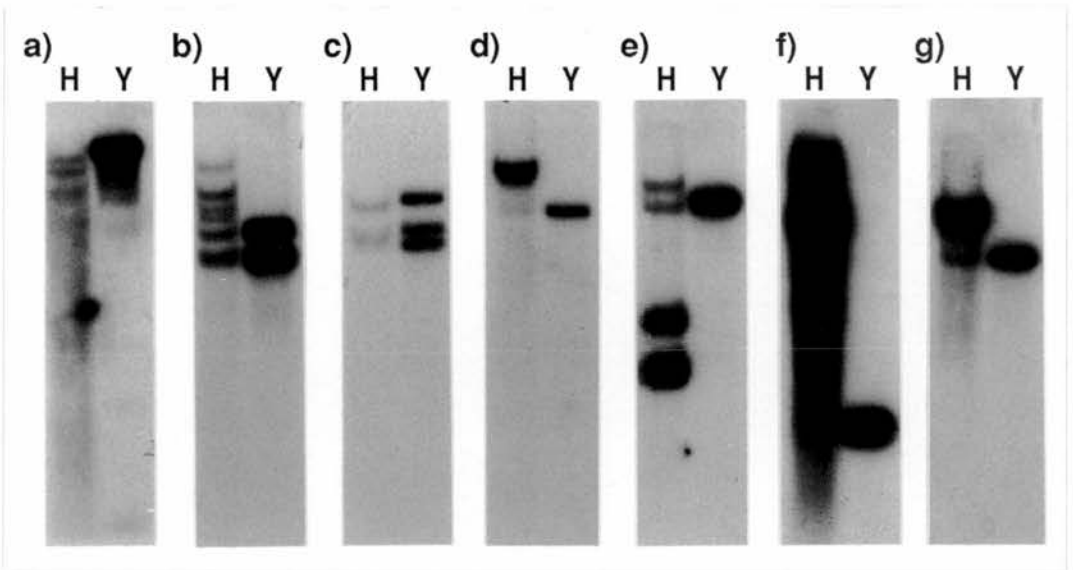


Figure 6.4 Hybridisation of the CSC novel cDNA fragments to *EcoRI*-digested total human genomic DNA (H) and *EcoRI*-digested total yeast plus YAC D0485 DNA (Y). a = nonran4, b = 2-10b, c = 5-3c, d = 4-7d, e = 1-11d, f = 4-1f and g = 3-3a (not aligned horizontally relative to each other). The results have been discussed in detail in the text.

6.4 Isolation of Cosmids Positive for the cDNA Fragments

The seven probes nonran4, 2-10b, 5-3c, 4-7d, 1-11d, 4-1f and 3-3a were used as hybridisation probes onto filters containing a gridded cosmid library of 520 clones (made by Yoshiro Shibasaki and Sheila Christie). Nonran4 gave a strong positive signal on 42 cosmids, 2-10b on 73, and 5-3c on 55. 4-7d, 1-11d, 4-1f and 3-3a hybridised to a set of 23 cosmids, with very few differences between them. A total of 520 cosmids of approximately 40kb each derived from a YAC of 1.3Mb should give a 16-times depth of coverage of the YAC. Thus for the probes nonran4 and 2-10b, which are present twice within the YAC, there should have been approximately 32 positive cosmids. The fact that both these probes hybridised to more cosmids than expected, especially 2-10b, suggests that the probes may contain a repeat sequence which occurs within, and is possibly specific to, the D0485 region. Alternatively some of the hybridisation signals may be false positives. A total of 48 positive cosmids would be expected for probe 5-3c, since it is present in three copies on the YAC. This is not very different from the actual figure observed of 55 positives.

The four probes 4-7d, 1-11d, 4-1f and 3-3a hybridised largely to the same set of cosmids. Figure 6.5 shows the results obtained from probes 4-7d and 3-3a. This implies that these four probes lie in close proximity (within ~40kb) on genomic DNA. One cosmid was positive for 4-7d, 1-11d and 3-3a only (c402), two for 4-7d, 4-1f and 3-3a (c564 and 568) and one for 4-1f alone (c620). Assuming these variations in hybridisation pattern were not caused by deletions, chimaerism or other aberrations in the cosmids themselves, the genomic order of the probes could be partly determined, (Fig. 6.6) although the orientation was not known.

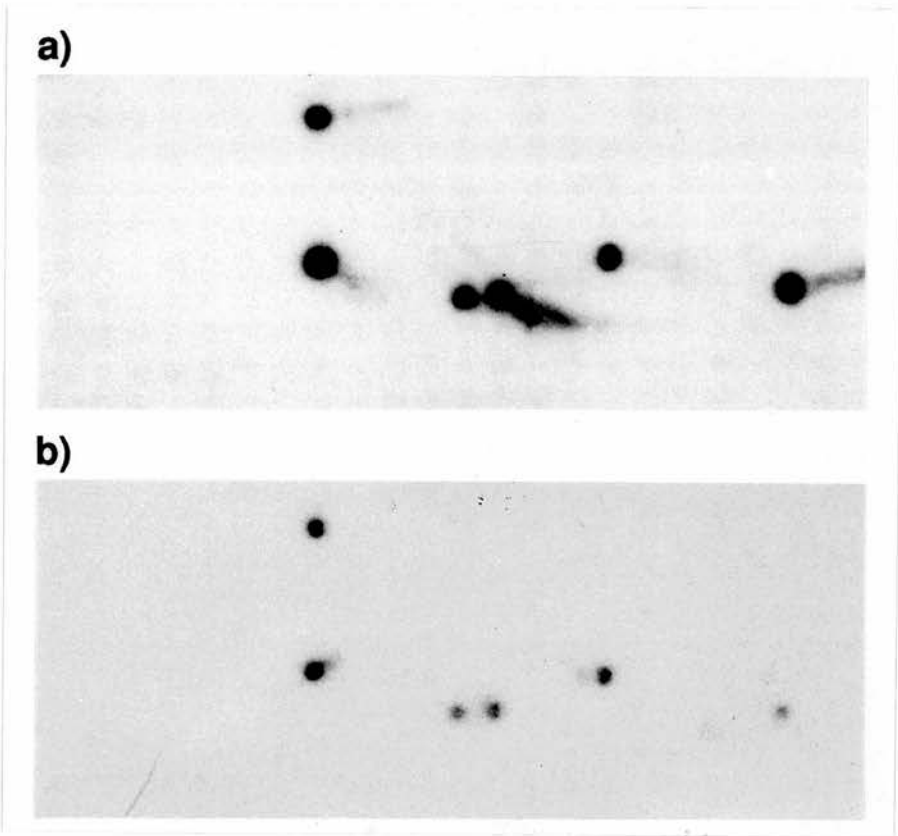
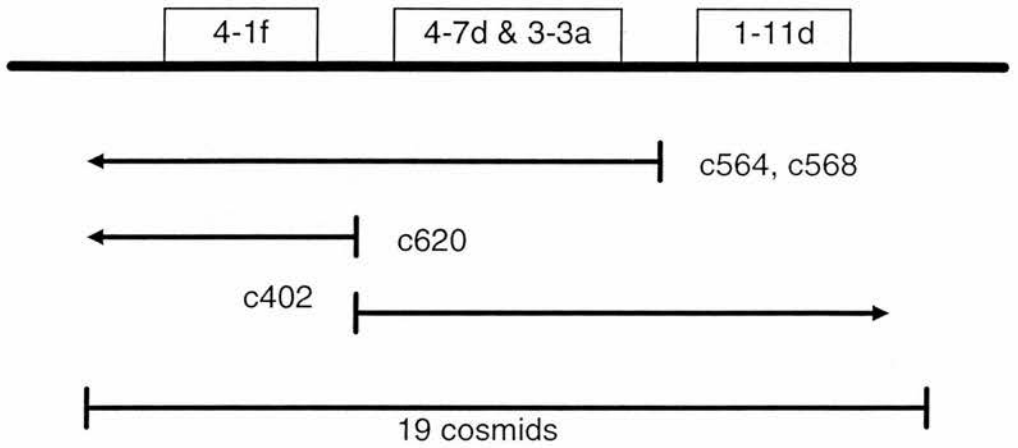


Figure 6.5 Hybridisation of the CSC novel cDNA fragment probes 4-7d (upper) and 3-3a (lower) to a gridded filter containing the D0485-derived cosmids 1 to 234 (approximately half of the total set). Both probes hybridise to the same six cosmids, which indicates that they lie in close proximity in genomic DNA.

Figure 6.6 Genomic order of probes 4-7d, 1-11d, 4-1f and 3-3a as defined by hybridisation to individual cosmids. The position of the ends of the cosmids is not known.

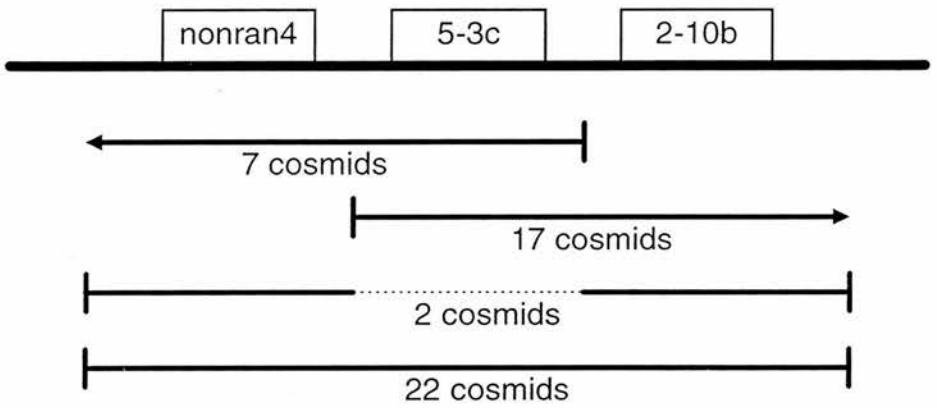


According to the pulsed field mapping data for the cDNA fragments (Fig. 6.2), there is no overlap between any of nonran4, 2-10b and 5-3c and the set of four fragments 4-7d, 1-11d, 4-1f and 3-3a. This observation is consistent with the results obtained from hybridisation to cosmids - there was no overlap in the set of cosmids positive for 4-7d, 1-11d, 4-1f and 3-3a with those positive for nonran4, 2-10b and 5-3c. (One cosmid, c157, which appeared to be positive for 2-10b and the set of four probes, was later shown by PCR to be a false positive for the group of four fragments, see table 6.3). There was however overlap seen in the cosmid sets identified as positives for probes nonran4, 2-10b and 5-3c. Two cosmids were positive for nonran4 and 2-10b, seven were positive for nonran4 and 5-3c, and seventeen were positive for 2-10b and 5-3c. Twenty-two cosmids were positive for all three of these probes.

For the three cDNA fragments nonran4, 2-10b and 5-3c there were cosmids positive for two of them in all three combinations. This is inconsistent with a linear genomic order for these sequences without implying false hybridisation data or deletions or other rearrangements in some cosmids. Since there

were only two cosmids positive for nonran4 and 2-10b alone, it may be most likely that these are the false positives or aberrant cosmids, and that the genomic order of the probes is as in Figure 6.7. The fine mapping of these probes would be complicated by probable low copy repeat elements within the probes, and the inverted repeat structure in the YAC, and was not attempted.

Figure 6.7 Most probable genomic order of probes nonran4, 2-10b and 5-3c, as judged by hybridisation to cosmids alone. A necessary deletion or other aberration in two of the cosmids is depicted by a dotted line. The orientation of the set is not known.



6.5 Fine Mapping of Fragments 4-7d, 1-11d, 4-1f and 3-3a

Since these four cDNA fragments all mapped in close proximity to each other, and the genomic locus was unfettered by complicating factors such as the YAC inverted repeat, a fine scale mapping of a cosmid positive for these four fragments was undertaken. Cosmid 158 was chosen, since it had been previously demonstrated (by Sheila Christie) that this cosmid contained a *Not I* restriction site.

Further restriction digests were performed on c158, to confirm the presence of the *Not I* site, and to investigate whether any other rare-cutter enzymes

(indicative of CpG islands) had sites within the cosmid. In addition, digests were also performed with the restriction enzymes used for the pulsed field mapping. Figure 6.8 shows the results of digesting c158 with the enzymes *Not I*, *Sal I*, *Sfi I*, *Sst II*, *BssH II* and *Eag I*. It can be seen that there are restriction sites for the enzymes *Not I*, *BssH II* and *Eag I* within this cosmid, which would be diagnostic for the presence of a CpG island if they occurred in close proximity to each other.

In order to map c158 more finely, it was necessary to digest the cosmid with restriction enzymes which cut DNA more frequently than those above, such that a detailed restriction map could be constructed. The enzymes *BamHI*, *Sma I*, *Sac I* and *EcoRI* were chosen for this purpose. Single and double digests of c158 were performed using these enzymes, plus those rare-cutter enzymes which had been found to digest the cosmid. Filters were made by Southern blotting of the restriction digest fragments which had been optimally resolved on long 0.8% agarose gels.

The four cDNA fragments (4-7d, 1-11d, 4-1f and 3-3a) were used as hybridisation probes onto these filters. All the probes gave positive signals, indicating that this cosmid was truly positive for these four fragments. Figure 6.9 shows the results generated with probes 4-7d and 3-3a. The sizes of the restriction fragments to which each probe hybridised was noted (see Table 6.2), which enabled the construction of a restriction map of the region of the cosmid containing the four probes (Fig. 6.10). Additional mapping data was generated by strategic excision of restriction fragments from gels, and then use of them as hybridisation probes. Combining the hybridisation data with knowledge of the restriction enzyme sites present in the cosmid vector (supercos; Evans *et al*, 1989) enabled the rare cutter sites to be physically linked to the cDNA fragments (Fig 6.10). Co-migrating and unresolvable fragments presented difficulties in mapping the extremities of the cosmid, but this was deemed unnecessary and was not pursued.

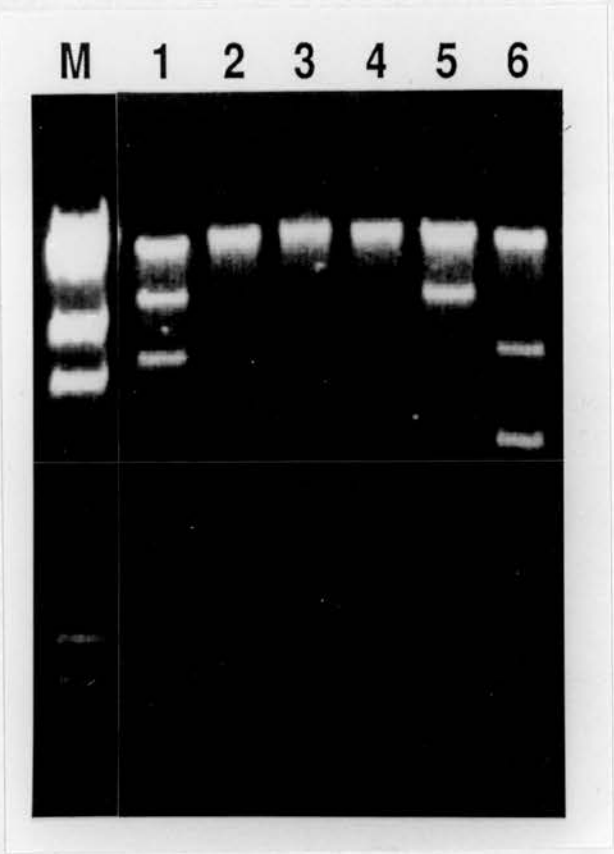


Figure 6.8 Products of digestion of cosmid 158 with rare cutter restriction enzymes, resolved on a 0.8% agarose gel. The size marker used is 250ng of λ*Hind III*. Track 1 = *Not I*, track 2 = *Sal I*, track 3 = *Sfi I*, track 4 = *Sst II*, track 5 = *BssH II* and track 6 = *Eag I*. The cosmid vector (supercos) contains two *Not I* sites and four *Eag I* sites; the human DNA insert therefore contains one *Not I* site, one *BssH II* site and two *Eag I* sites (in track 6 the third largest band is a doublet, and the two small vector-derived bands are not visible).

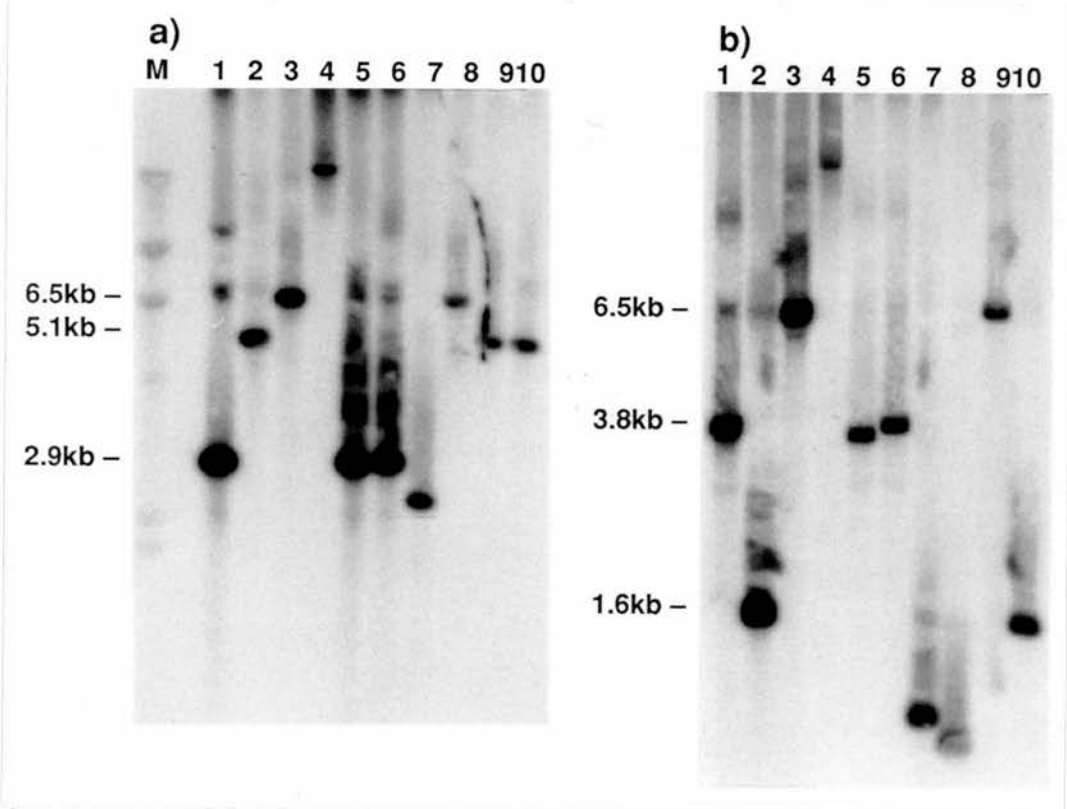


Figure 6.9 Hybridisation of the CSC novel cDNA fragment probes 4-7d (left) and 4-1f (right) to restriction digested cosmid 158. The marker track visible on the 4-7d autoradiograph is λ*Hind III*. Tracks 1 to 10 correspond to *BamHI*, *EcoRI*, *Sac I*, *Sma I*, *BamHI* & *Sac I*, *BamHI* & *Sma I*, *EcoRI* & *BamHI*, *Sac I* & *EcoRI*, *Sac I* & *Sma I* and *Sma I* & *EcoRI* digests respectively. The sizes of the hybridising bands (except *Sma I* single and double digests) are shown in table 6.2. Hybridisation to rare cutter restriction enzyme digests is not shown. The restriction map constructed partially from this data is shown in Fig. 6.10.

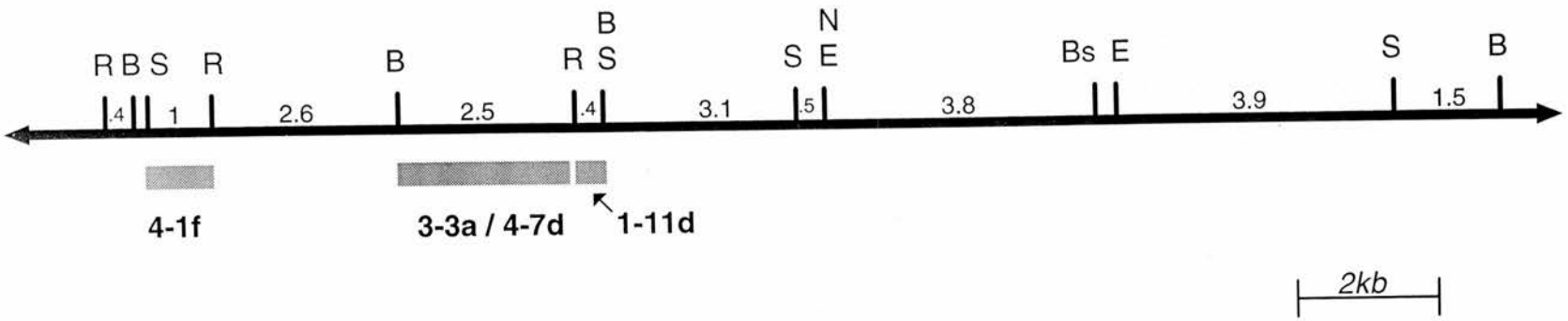
Table 6.2						
	Restriction fragment sizes to which the probes hybridised (kb)					
Probe	BamH I	Sac I	EcoR I	BamH I / Sac I	BamH I / EcoR I	Sac I / EcoR I
4-7d	2.9	6.5	5.1	2.9	2.5	5.1
1-11d	2.9	6.5	6.4	2.9	0.4	0.4
4-1f	3.8	6.5	1.6	3.6	1.2	1.0
3-3a	2.9	6.5	5.1	2.9	2.5	5.1

It can be seen from Fig. 6.10 that the four cDNA fragments, 4-7d, 1-11d, 4-1f and 3-3a all map to within a 6.5kb *Sac I* fragment. The nearest fragment, 1-11d, lies only 3.6kb away from a *Not I* and *Eag I* site, 7.4kb away from a *BssH II* site and 7.7kb away from a second *Eag I* site. According to the long range mapping data (Fig. 6.2), this must correspond to the distal-most *Not I* site on the YAC D0485, enabling the placement of the four cDNA fragments on the pulsed field map with far greater accuracy.

The order of the four cDNA fragments derived by hybridisation to the whole cosmid set was the same as that derived from the detailed mapping of cosmid 158. Probes 4-7d and 3-3a were both located within a 2.5kb *BamHI* & *EcoRI* fragment and their order could not be deduced by digesting the cosmid with the enzymes used. This was resolved by PCR (section 6.6.3).

This mapping data is suggestive that the four novel cDNA fragments 4-7d, 1-11d, 4-1f and 3-3a are all derived from a single, CpG island-associated gene. RT-PCR could be used to investigate whether these fragments are derived from the same transcript. Analysis of the methylation status of the rare-cutter restriction enzyme sites in uncloned genomic DNA will determine whether or not they mark a CpG island.

Figure 6.10 Restriction map of 20.2kb of cosmid 158. *BamHI* (B), *Sac I* (S), *EcoRI* (R), *Not I* (N), *BssH II* (Bs) and *Eag I* (E) sites are shown. The restriction fragments containing the four cDNA fragment probes 4-7d, 1-11d, 4-1f and 3-3a are indicated by grey bars. The numbers represent the distance between adjacent restriction sites, in kb.



6.6 Linking the cDNA Fragments by Genomic PCR Amplification

6.6.1 Primer Design

In order to confirm the order and separating distance of the group of four cDNA fragments, and to potentially link the other three fragments (nonran4, 2-10b and 5-3c), PCR primers were designed from the consensus sequences of the five families of cDNA products and the two individual products. Due to the alternatives in orientation available to the products, two primers in opposite orientations were required. These primers were designed such that it would be possible to amplify a short intra-fragment product, to serve as a positive control for the template DNA. Ambiguous bases were avoided where possible. Figure 6.11 A to G shows the consensus sequences of the seven cDNA fragment families and products, with the primers' name and orientation indicated for each one.

Figure 6.11

A) Family 1 - 482/483 PCR product=198bp

```
                482 ⟶
1    AGGAGAAGCT CGTGGCTGCT GAGTTCTCCT GGCCACCATG AACTTCAGGA
51   AGTGGGTGCT ATAGCAGCTG CCTGAACTAC ACAATCTGGG CTTTGGTGTA
101  TCCCTGTATG CCCTCCGGGC CAGACACTGG AGGTGTCATT TCCAAAGCAA
                                    ⟵ 483
151  ATTGGAAGCG CTTTTTTGGA ATTTCTCTCC AATGCTTTCT ACTCACAAAG
201  ACTGACATCT TAACACGTGG CAAAGAAAAA AATATTTAAA GGGTCCAGAT
251  CTTATTTATG TAAACAATCA AGAGTGAGTT TGKRGKKGRA AACCCAAAGT
301  TGGANAANTG GTGCATAATA AAAGAAATTT GTTCATTTTC TGCTGTTGTA
351  CATTTGGAAT GATTTTTGTA TTTTGATTTT GTGAACATGT CTCCTAATGT
401  AAATAGCCAA TAGATTCTCT TTTCGGACAG TATCTTCTGA TAGCTGGAAT
451  GTCTGGGTTA TAAAATTTGT GATC
```

B) Family 2 - 484/485 PCR product =200bp

```
1    GATCTGGTGG AAACACAGGC TGCCAGGCCC TGCCCCTGGA GTCTTTCTTT
                                    484 ⟶
51   GAATAGGCCT GGGGTGGGGC CCAAGAATGA GCAGAAGAAC AGGTTTCCTG
101  GTGAGGCTGA TGCCGCTGGC CCAGGGTTCT CACGTTGAGG ACCGAGAGCT
151  TTGAGGTTTT CATACCTGAT GATGTCCAGG AACCCTTCTC AGTAGGCACT
201  GAAGACCATC AGCAGAATCA CAGACCCCAG GAGAGATGTC GTCAGACAGA
          ⟵ 485
251  CACAGAGGCA TCACCGAATT AAAGTGAAAA TGAAGAAAGG AGCTGAGCAT
301  CTGTTTCATG ACTTTCGTGG CTGTTTTACT AAAGAGGCTA TCCTGGCCCA
351  GTCAGGGCAC GCTATCATCA CCAACTACTT GTTGAACTAT GTCCTGGGTC
401  TTGACCTTGA AGGATGGAHG CGTTBGGGTT GCAAGCCGCC ATGCAGGGTC
451  GAACCGATCC
```

C) Family 3 - 486/487 PCR product =153bp

```
                                    486 ⟶
1    GATCCGTGGG AGCACTGAGA TGGTCATCTC CAAATATCAC ATATCTCACA
51   TCCTCAGAAA GGCTTATATA GCAAGGAATC ATTTCCGTGC TCAGAGCACT
101  ATCCACTCTG AAGTTGTTGA GCATGTCTAG GACTCCAGTT ATGCACCATG
                    ⟵ 487
151  AAGTGAGCTC TGGGTTCACT GGCTGGGGCT TTTGSATCTG TGCCAAATCA
201  GTCCTTGCTG ATACATTTCT CACATCCTTG
```

D) Family 4 - 488/489 PCR product =255bp

```
                488 ⟶
1    GATCCAAATG GCATCTCCTA CCCTATCCCT GTATGGCTTA ACCTAGGGAA
51   ACTCTTTCCC AGAGGAGTCA GGTAAGGAGA CGGTGGCTGA GCTTCTTACA
101  GACTTAAAGG AGGCATCCTG GAATTTRGGA SWCCRTCCTK CCWTTCTGTA
151  AYCMCTGGCA GCTSCTGCTG CTKCKCARAG TTTAGCTTTG TCTCTCAWCC
                                    ⟵ 489
201  AGCTCAGACT GKTGCTGGTC CTGATGGCCT CTGCWTAGCT GTATTAGATG
251  TCCTTGCRTT GCWAAAAAGA AATACATGAG ARATACCTTT AGKTGGCTGA
301  TGGTWCCATA GGGDATACAG GAAGCRTGGT GGTTDCTGCT TGTGGGGAGG
351  CCTCAGGAAA CTTTTACTCA AGGTACAGGC CCAATGAGKA AAAAGAAAAA
401  AAAAAAAGCT ACCAGCRCCG CAATGTGGAG TCCTGAKC
```

E) Family 5 - 490/491 PCR product =239bp

```
1    GAATTCTATG CCTCAGGTTY TCATTTTTGT TATGTGCAAA ATAAACTCGT
                                              490 →
51   ATCCTGCTAT AGATYTAAAC CTGTGAGVGA MTCCTACTCA AAACCTGATT
101  CAAACTTGTG TGTAGATCTT TGYCTCTCTG TGGCATAMRK AATMAWTCCT
151  TCTGGTTTMC TTCCCTCAGA ATAATGGACT TAGACTKCCC ACAAGMCTTC
201  CRGAYAGWMM YCACCTGKCT TATCTGSCTG ARTTACCTCA TAGACCCCAT
251  CACTATAGGC TGSGGGGACA GCYTCTGTAG GCCCTGCCTC TGCCTTMGCT
                  ← 491
301  GGGAAGAAGY ACACACTCCT GCCCYGCATG CAGGGAATCG TCACAGCAGG
351  AAGATTTC
```

F) Product 6 (4-1f) - 492/493 PCR product =107bp

```
1    CGTGAGCCAC CACCCCCGGC CCCAGGACAC ACAGCTTTAA AATTTCTCCT
                                         492 →
51   TGGTCTCACC CAGTGCCAAC CACCTAAAAC CTCTCATTTT CCCCCAGACA
101  TTTCTTCTGC CTCCAGGATG GAGGTAGAGA ATCTTGGCCT AGGCCCACGC
                  ← 493
151  ACTGGGGACC ATGCTGGGCT GCCGTGGACA GTGACGGACT CAGGTTCTCA
201  CCAGGATC
```

G) Product 7 (3-3a) - 494/495 PCR product =264bp

```
1    GATCTAGGAG CAGAGGGCAG AGCCTCAGCA GGAAGAGCGT CTCTTTGAGA
51   AGGAGACACA GTGGAGCAGG TGTGTAGGTT CACAGGGCCA GCTATGGGTA
101  GAGTCGGGTG TACATTTTTA GGAGCCACAA TTCCCAAAAA TCTCCTGACT
151  ATAACATCAG TGCACAGAGC CAGTCAAATG GAGGAGGAGT GGGTCCAGGC
              494 →
201  AATTCAGGAA GAAGGAAAGT AACAAATGAG TGGTTGCAGG AGGACACTTT
251  TTCTTGTCGA GGGTCACTAA ACAAAACATT GTCTCCTCCC CCTCTAACTT
301  CAGAAACAAT GGAGGGTAAG AGTGTCNCCT GGGCCCTGGG GCCAAAGACA
351  GTAGATAACT TCTCTGTCGT GTTCTCCAGA AGGGCCCAAC AANTACAAGG
401  TTCTACGGTT CTAAATTCCA ATCTAGTCTT CCACATCATT TTGAAGGTAT
                                  ← 495
451  AATATTACTT GTCAAAGTGG GATGATAGAA GATATGTGTG GACATAAATT
501  GTTGTCAA
```

6.6.2 Intra-fragment PCRs

Several of the cosmids which had been positive by hybridisation for one or more of the cDNA fragments were picked as templates for intra-fragment PCR with the primer pairs derived from the same cDNA fragments. This was to test whether the cosmids were positive for the expected fragments, before using them in inter-fragment PCRs.

The cosmid colony was picked from a fresh agar plate, and transferred to a microfuge tube containing 50µl of dH_20. The tube was the heated to 96°C for 5 mins, the contents spun down for 1 min to separate the DNA solution from the bacterial cell debris, and then placed on ice. 2µl of template DNA plus 300ng of each primer was used in a PCR reaction. For the intra-fragment PCRs the programme used was:

Hot start at 90°C for 5 mins

Denaturation: 94°C for 30 secs in the first cycle and 15 secs thereafter

Annealing: Touch down from 70°C to 60°C then 60°C, for 30 secs

Extension: 72°C for 30 secs (cycles 1-10), 1 min (11-20), 1.5 mins (21-30)

It was found that not all the cosmids which had been positive for a product by hybridisation were also positive by PCR. Table 6.3 summarises these results. There are several alternative explanations for this phenomenon. Firstly, the hybridisation results could have been false positives. Secondly, deletions or other aberrations in the cosmids could have removed one or both of the primer binding sites, yet retained sufficient of the cDNA fragment sequence that hybridisation could still take place. Thirdly, intact cosmids could be positive by hybridisation but negative by PCR if the probe cDNA sequence was located at the extreme terminus of the cosmid insert. The PCR primers were derived from the consensus sequence of the cDNA families, and therefore sometimes lay outwith the sequence of the representative family member which was used as a probe (that is, nonran4, 2-10b, 5-3c, 4-7d, 1-11d, 4-1f and 3-3a).

Table 6.3 Fam = Family; ND = not determined		
Cosmid	**Expected PCR Positives**	**Observed PCR Positives**
3	Fams 1,2 and 3	Not Fam 1, Fams 2 & 3 ND
22	Fam 2	Fam 2
24	Fam 1	no positives
26	Fams 2 & 3	no positives
42	Fams 1 & 3	Fams 1 & 3
44	Fams 1 & 3	Fams 1 & 3
53	Fam 3	Fam 3
85	Fams 2 & 3	Fams 2 & 3
129	Fams 4, 5, 6 & 7	Fams 4, 6 & 7
157	Fams 2, 4, 5, 6 & 7	Fam 2
158	Fams 4, 5, 6 & 7	Fams 4, 6 & 7
208	Fams 1 & 2	no positives
499	Fams 1 & 2	Fam 2

No PCR product could be obtained from Family 5, even though its existence on at least c158 had been proven by the mapping of the cosmid. The failure could therefore be ascribed to the PCR reaction. Primer 491 was later found to function adequately in the inter-fragment PCRs, and thus it is probable that the problem was caused by primer 490. No optimisation of PCR conditions was attempted in order to resolve this.

Neither of the cosmids (208 and 499) which had been positive for Families 1 and 2 by hybridisation were found to be also positive for these fragments by PCR. This may resolve the discrepancy shown in Fig. 6.7.

For Families 1, 2, 4 and 6, the band obtained by intra-fragment PCR was the same size as that expected, that is 198bp, 200bp, 255bp and 107bp respectively. For Families 3 and 7 however, the product band was larger

than expected. The band in Family 3 was approximately 1.3kb rather than 153bp, and the band in Family 7 was approximately 500bp rather than 264bp. This implies either non-specific annealing of one or both of the primers at a more distant site, or the presence of introns within the genomic DNA. Sequence analysis was undertaken to investigate these possibilities (see Section 6.7).

6.6.3 Inter-fragment PCRs

The cosmids which had been proven to be positive for two or more of the cDNA fragment families could be used as a template for PCR reactions attempting to link the different families in genomic DNA. Due to the alternative orientations and orders available to the fragments all four combinations of the two primer pairs were required, only one of which should result in a product. The order of the fragments comprising Families 4/7, 5 and 6 had been previously established by the mapping of cosmid 158, which also established the approximate distance separating them. Results from hybridisation of probes nonran4, 2-10b and 5-3c onto the cosmid set had indicated a likely order for these fragments, but nothing was known regarding the distance separating them.

A long PCR kit was employed ('Expand'™ or 'Takara'™) for many of the reactions, in order to increase the likelihood that a PCR product could be generated. 2µl of cosmid template (prepared as in Section 6.6.2) was used in the majority of the PCR reactions, although in those indicated a product was also obtained from 1µl (~100ng) of YAC D0485 miniprep DNA. 300ng of each primer was used; other components of the reaction mix were according to the manufacturer's instructions. The PCR programmes used differed only in the extension times, adjusted according to the length of the target product:

Hot start at 90°C for 5 mins

Denaturation: 94°C for 30 secs in the first cycle and 15 secs thereafter

Annealing: Touch down from 68°C to 58°C then 58°C, for 30 secs
Extension: 68°C, 5 x 3 mins, then 25 x 6 mins[1] or 15 mins [2].

Where the long PCR kit was not employed [3], the PCR programme was:
Hot start at 90°C for 5 mins
Denaturation: 94°C for 30 secs in the first cycle and 15 secs thereafter
Annealing: Touch down from 70°C to 60°C then 60°C, for 30 secs
Extension: 72°C for 2 mins (cycles 1-10), 4 mins (11-20) and 6 mins (21-30).

Table 6.4 shows the results of the inter-fragment PCR reactions, with the cosmid template used, the combination of primers used for successful amplification of the target product, the size of that product, and the PCR programme used ('Expand'™ or 'Takara'™, (1) , (2) or standard Taq polymerase (3), see above text). No product was obtained in each case with any of the remaining three combinations of primers.

Table 6.4

Product	Cosmid template	Primer Combination	Size (kb)	Programme used
Fam 1 - Fam 3	44	482 / 486	12.0	Takara (2)*
Fam 2 - Fam 3	85	485 / 487	6.5	Takara (2)
Fam 4 -Fam 5	129 & 158	488 / 491	2.4	(3)
Fam 4 - Fam 6	↓	489 / 493	4.1	Takara (2)
Fam 4 - Fam 7	↓	489 / 495	0.9	Expand (1)
Fam 5 - Fam 6	↓	491 / 493	6.4	Takara (2)
Fam 5 - Fam 7	↓	491 / 495	3.1	Expand (1)

* = also obtained with Expand (2)

For the PCR reactions Fam 1 to 3, 2 to 3, 4 to 6, 4 to 7 and 5 to 6, a product was also obtained when YAC D0485 was used as a template. This was not attempted for the other two products.

Figure 6.12A shows the PCR products generated from cosmid DNA between Families 1 and 3, and 2 and 3. No product was obtained between Families 1 and 2. Fig. 6.12B shows the products similarly obtained between Families 4 and 5, 4 and 7, 4 and 6, 5 and 6, and 5 and 7.

The combined results from the inter-fragment PCR reactions enabled clarification of the order and orientation of all the seven group of cDNA fragments (Figure 6.13). The fragments fall into two groups, that comprising Families 1, 2 and 3, and that comprising Families 4, 5, 6 and 7. This confirms speculation arising from the pulsed field mapping data and from the hybridisation to the cosmid set.

PCR reactions were attempted to amplify between the two sets of fragments on genomic DNA. Knowledge of the order and orientation of the fragments enabled the primers to be chosen so as to minimise the length of the target product (that is, 483 or 484 paired with 490 or 492; see Fig. 6.13). The simplest template DNA for these reactions was however YAC D0485, since no cosmids had been identified which contained fragments from both groups. No PCR products were generated, even with use of the 'Expand' long PCR kit. This is unsurprising when mapping position of the two groups on the YAC D0485 is considered. The location of the first group (Families 1, 2 and 3) is fixed by the *Sfi I* site present between Families 2 and 3 (see Fig. 6.2). The location of the second group (Families 4, 5, 6 and 7) is also fixed, by its close proximity to a *Not I* site demonstrated by the mapping of cosmid 158. The distance between the *Sfi I* site and *Not I* referred to is 170kb, which is not amplifiable by PCR (even using a kit!).

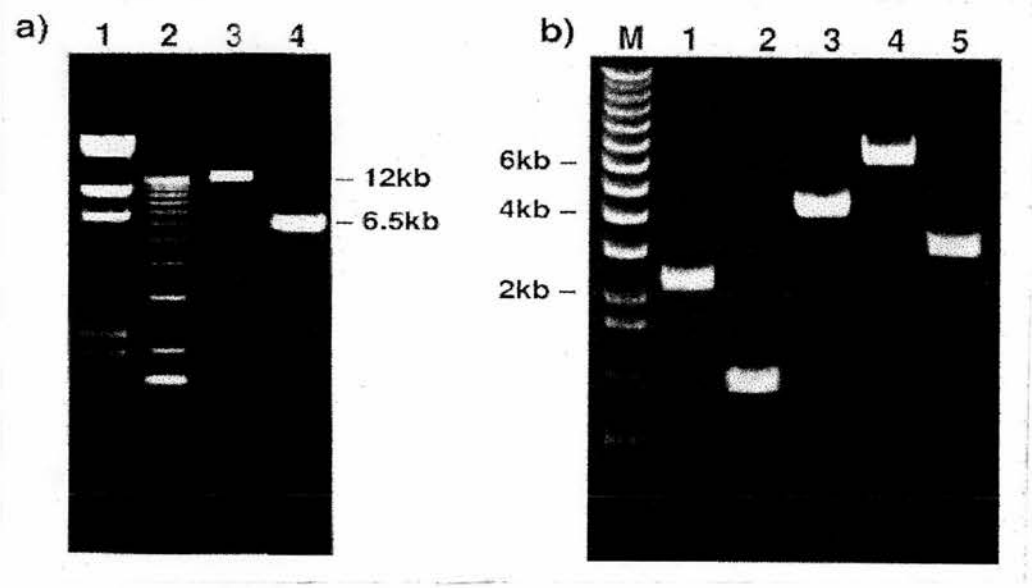
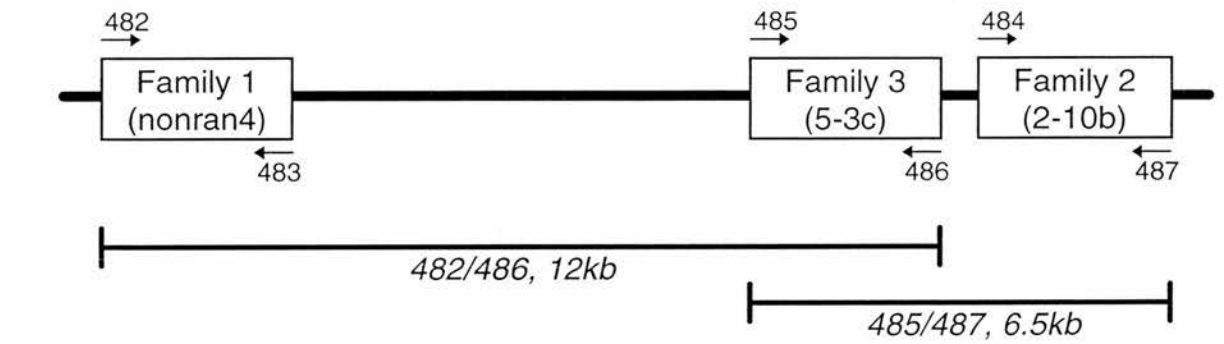


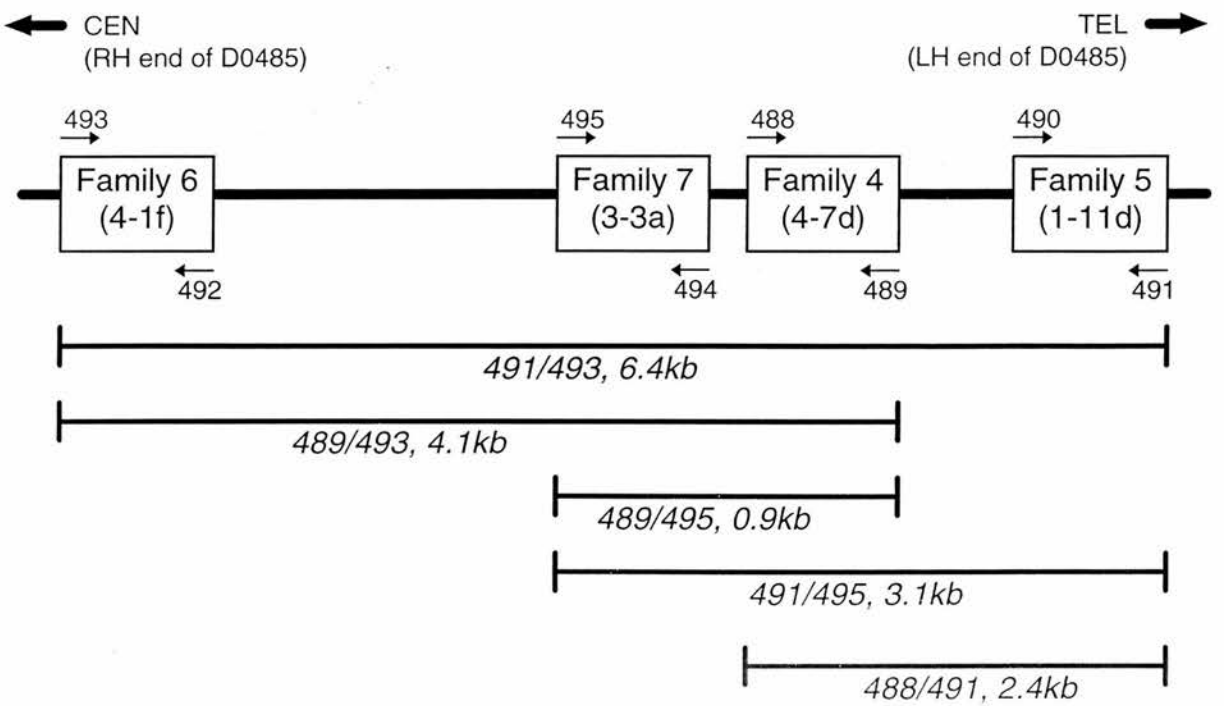
Figure 6.12 Inter-fragment PCR products generated by amplification of cosmid DNA with primers designed from the novel cDNA family consensus sequences (see table 6.4). 10µl PCR product is resolved on a 0.8% (a) or 1% (b) agarose gel.

a) The size markers used are 250ng λ*Hind III* (track 1) and 250ng of the 1kb ladder (track 2). Track 3 = 12 kb fam1 to fam3 product. Track 4 = 6.5 kb fam2 to fam3 product.

b) The size marker used is 250ng of the 1kb ladder. Track 1 = 2.4 kb fam4 to fam5 product. Track 2 = 0.9 kb fam4 to fam7 product. Track 3 = 4.1 kb fam4 to fam6 product. Track 4 = 6.4 kb fam5 to fam6 product. Track 5 = 3.1 kb fam5 to fam7 product.

Figure 6.13 Order, orientation and genomic separation distance of cDNA fragments within the two groups, as defined by PCR (not to scale). The first set probably occurs in both orientations on the YAC, due to the inverted repeat structure. The second set probably occurs in the order shown, since 1-11d was found to map closest to the *Not I* site on cosmid 158. CEN = centromere, TEL = telomere.





6.7 Sequencing of the YAC D0485 Copy of the cDNA Fragments

All of the seven novel cDNA fragments, except 5-3c, were found by hybridisation to total genomic DNA to exist at loci in the human genome other than within YAC D0485. The selection procedure used to isolate the coincident DNA will enrich for all the cDNA fragments which form a sufficiently stable duplex with the genomic DNA. As such it was uncertain whether the cDNA fragments isolated were derived from transcription of a gene on YAC D0485 itself, or from a very similar gene transcribed from elsewhere in the genome. The cDNA fragments were certainly sufficiently similar to the D0485 copy to give strong clear hybridisation signals on the YAC and associated cosmids, and only genomic sequence analysis could reveal the single base changes permitting the assignment of the cDNA to D0485 or to 'elsewhere in the genome'.

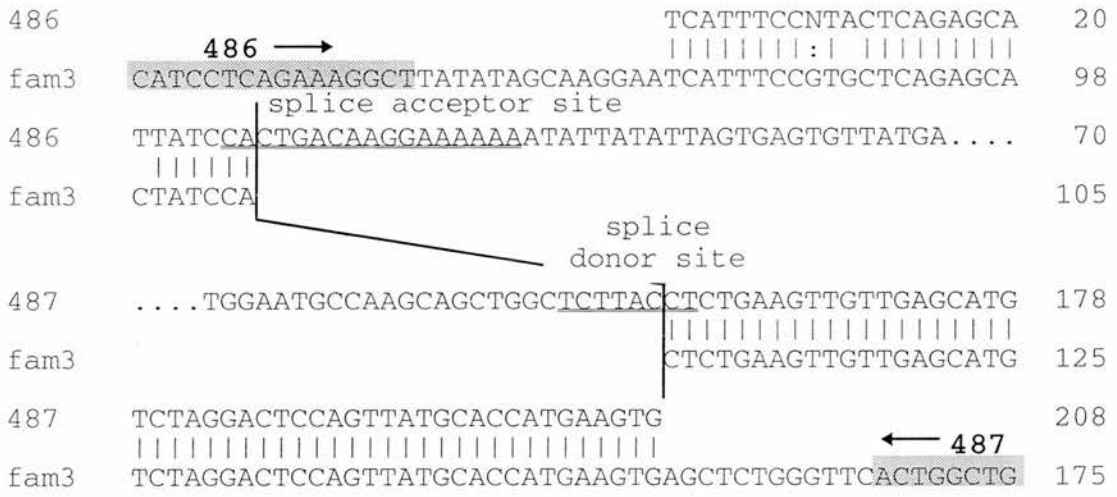
Sequencing of the genomic copy of the cDNA fragments was also important to verify that the intra- and inter- fragment PCR products were genuinely amplified from the correct fragments. Thirdly, it was important to ascertain whether the larger than expected intra-fragment PCR products generated from Families 3 and 7 were as a result of intervening introns.

The ends of selected inter-fragment PCR products were sequenced using dideoxy terminator radioactive sequencing, from the primers used to amplify them. No sequence data was generated for either Family 1 or Family 2. Sequence was generated for Family 3 from primers 486 and 487 (using the 1.3kb intra-fragment PCR product as a template), for Family 4 from primers 488 and 489 (using the inter-fragment PCR products 488/491 and 489/493), for Family 5 from 491 (using 488/491), for Family 6 from 493 (using 491/493 and 489/493) and for Family 7 from 494 and 495 (using 491/493 and 491/495). Approximately 200bp of sequence data was generated from each primer.

The sequences obtained from the D0485 copy of the cDNA fragments were compared with those derived from the cDNA fragments themselves. Figure 6.14 A to E shows a comparison of the sequence of the genomic and cDNA copies of each of the Families 3 to 7.

Figure 6.14

A) Genomic sequence generated from primers 486 and 487 compared with the Family 3 cDNA consensus sequence.



The genomic sequence diverges from and then re-converges with the cDNA sequence at position 106 of the Family 3 cDNA consensus sequence, which suggests the presence of an intron. The presence of an intron would be consistent with the intra-fragment PCR product from genomic DNA which was much larger than expected. In proof of this, reversed consensus splice donor and acceptor sites (Shapiro and Senapathy, 1987) can be found at these junctions (bases underlined). The acceptor site sequence has 2 bases additional to the consensus (YYTTYYYYYYNC/AGG) and the donor site has 7 out of 8 bases identical to the consensus (AG/GTRAGT). These differences could have arisen from sequencing errors, or be the result of mutations generating cryptic splice sites.

Throughout the 77 bases of matched exon region between the genomic and cDNA sequence there are only two mismatches, which may be attributable to sequencing errors. The cDNAs comprising Family 3 were therefore all derived from a D0485 YAC copy of this sequence. This is consistent with the total genomic hybridisation data for 5-3c, which suggested that there were no genomic copies of this sequence other than within D0485.

B) Genomic sequence generated from primers 488 and 489 compared with the Family 4 consensus sequence.

```
488                                          GGAAACTCTTTCCCAGAGGA   20
          488 ⟶                              ||||||||||||||||||||
fam4      CCTACCCTATCCCTGTATGGCTTAACCTAGGGAAACTCTTTCCCAGAGGA   66

488       GTCAGGTAAAAAGACAGTGGCNGAGCTTCTTACAGATTTAAAGGAGACAT   70
          |||||||||  |||| |||||:|||||||||||||| |||||||| |||
fam4      GTCAGGTAAGGAGACGGTGGCTGAGCTTCTTACAGACTTAAAGGAGGCAT  116

488       CCTGGAATTTAGGACTCCATCCTTCCTT                         98
          ||||||||||:|||::||:||||:||:|
fam4      CCTGGAATTTRGGASWCCRTCCTKCCWTTCTGTAAYCMCTGGCAGCTSCT  166


489                                ATCCC-GTATGGCTTAACCTAGGGAA   25
                                   |||||  ||||||||||||||||||||
fam4      GATCCAAATGGCATCTCCTACCCTATCCCTGTATGGCTTAACCTAGGGAA   50

489       ACTCTTTCCCAGAGGAGTCAGGTAAAAAGACNGTGGCTGAGCTTCTTACA   75
          |||||||||||||||||||||||||  ||||:||||||||||||||||||
fam4      ACTCTTTCCCAGAGGAGTCAGGTAAGGAGACGGTGGCTGAGCTTCTTACA  100

489       GATTTAAAGGAGACATCCTGGAATTTAGGACTCCATCCTTCCNTTCTGT-  124
          ||  |||||||||  |||||||||||||:|||::||:||||:||:||||||
fam4      GACTTAAAGGAGGCATCCTGGAATTTRGGASWCCRTCCTKCCWTTCTGTA  150

489       -CCCCTGGCAGCTCCTGCTGCTTCTCAAAGTTTAGCTTTGTCTCTCATCC  173
           :|:|||||||||:|||||||:|:||:||||||||||||||||||:||
fam4      AYCMCTGGCAGCTSCTGCTGCTKCKCARAGTTTAGCTTTGTCTCTCAWCC  200

489       AGCTCAGACTGTTGCTGGTCCT                               195
          ||||||||||:|||||||||                  ⟵ 489
fam4      AGCTCAGACTGKTGCTGGTCCTGATGGCCTCTGCWTAGCTGTATTAGATG  250
```

A total of 197 bases have been sequenced from the genomic region corresponding to the Family 4 sequence, 90 of these also sequenced in the opposite direction. Of these 197, there were 8 different bases between the genomic and cDNA sequences, 5 mismatches (although one was ambiguous), and three bases deleted from the genomic sequence. The Family 4 was composed of sequence from two clones, 4-7d and 5-7e, and where they differed, a degenerate base was inserted into the consensus. At all the positions of these degenerate bases, the nucleotide in genomic DNA was compatible with the degenerate cDNA nucleotide. At ten of these positions, the base in genomic DNA was the same as the sequence from 4-7d. However at four degenerate positions and at two of the three deleted bases, the genomic DNA base was the same as that in 5-7e. Some of these mismatches will undoubtedly be sequencing errors, whereas some may reflect genuine sequence changes between different copies of this sequence throughout the genome. It is therefore difficult to ascertain whether either or both clones 4-7d and 5-7e were derived from the D0485 copy of the gene, although they are highly similar.

C) Genomic sequence generated from primer 491 compared with the Family 5 consensus sequence

```
491                                          CTTTGCATAGATCTTTGTCTC   21
                                               |  ||  ||||||||||:|||
fam5   GVGAMTCCTACTCAAAACCTGATTCAAACTTGTGTGTAGATCTTTGYCTC  126

491    TCTATGGCATA-AGAATAAATTCTTCTGGGTTTCTTCCCTCAGAAAAATG   70
       |||  |||||||  ::|||:|:|  |||||||  ||  ||||||||||||  ||||
fam5   TCTGTGGCATAMRKAATMAWTCCTTCTGGTTTMCTTCCCTCAGAATAATG  176

491    GACTTAGAATTCCCACAAGCCTTCCAGAAAGAACTCACCTGCCTNATCTG  120
       ||||||||  |:|||||||||:||||||:||  ||::::||||||  ||:|||||
fam5   GACTTAGACTKCCCACAAGMCTTCCRGAYAGWMMYCACCTGKCTTATCTG  226

491    CCTGAATTACCTCATAGACCCNATCACTATAGGCTGCTGGCACAGTTTCT  170
       :||||:||||||||||||||||:||||||||||||||:  ||  ||||  :|||
fam5   SCTGARTTACCTCATAGACCCCATCACTATAGGCTGSGGGGACAGCYTCT  276

491    GTAGGCCCTGC                                         181
       ||||||||||                    ←— 491
fam5   GTAGGCCCTGCCTCTGCCTTMGCTGGGAAGAAGYACACACTCCTGCCCYG  326
```

181 bases of the Family 5 sequence have been sequenced from genomic DNA, 16 of which are different from the Family 5 consensus (15 mismatched bases and 1 deletion). This 9% difference rate is higher than would be usually attributable to sequencing errors and may therefore suggest that the cDNA clones comprising the Family 5 consensus (1-11d and 2-5c) may not be derived from the D0485 copy of the sequence. At the corresponding genomic DNA positions of the 16 degenerate bases in the Family 5 consensus, 7 were the same as 2-5c and 9 the same as 1-11d.

D) Genomic Sequence generated with primer 493 compared with the Product 6 (4-1f) sequence

```
493     CCTCGGCCTCTCAAGTGCTGGATTACAGGNGTGAGCNACCGCGCCCGACC   83
                                      |||||| :|||  |  ||||  ||
4-1f                                  CGTGAGCCACCACCCCCGGCC   21

493     CCAGGACACACAGCTTTAAAAGCTCTCCTTGGTCTCACCCAGTGCCNACC  133
        ||||||||||||||||||||  ||||||||||||||||||||||||:|||
4-1f    CCAGGACACACAGCTTTAAAATTTCTCCTTGGTCTCACCCAGTGCCAACC   71

493     AACTAAAACCTCTCATTTTCCCCCAGGCATTTNTTCTGCCTCCAGGATGG  183
        | ||||||||||||||||||||||||  |||||:|||||||||||||||||
4-1f    ACCTAAAACCTCTCATTTTCCCCCAGACATTTCTTCTGCCTCCAGGATGG  121

493     AGGTACAGNATCTTGGCCTTGGG                             206
        |||||  ||:||||||||||  ||                 <-- 493
4-1f    AGGTAGAGAATCTTGGCCTAGGCCCACGCACTGGGGACCATGCTGGGCTG  171
```

The genomic sequence generated from primer 493 extended further than the 4-1f sequence itself. This extra sequence may extend the known cDNA sequence, though it may instead continue into an intron. The former is more likely as there is no evidence for a splice site in this additional sequence. Of the 144 bases of the genomic DNA corresponding to the cDNA however, there were 9 definite mismatches and possibly 4 others in positions where the genomic base could not be determined. This 6% mismatch rate may be due to an unusually high rate of sequencing errors or may imply that the 4-1f cDNA was not derived from D0485 itself, but from a highly related gene elsewhere in the genome.

E) Genomic sequence generated from primers 494 and 495 compared with the Product 7 (3-3a) sequence

As with the Family 3 sequence, the genomic and cDNA Family 7 sequences do not match along their whole length. Rather, the genomic sequence diverges away from the cDNA sequence at cDNA position 389 and then reconverges at the same point. This is suggestive of the presence of an intron. The intra-fragment PCR product generated from genomic DNA was slightly larger than expected, which is consistent with the sequence data.

The Family 7 genomic sequence does not however exhibit consensus splice sites at the intron/exon boundaries. The 'GT' dinucleotide at the beginning of the mismatch (underlined) does suggest the 5' end of an intron (Shapiro and Senapathy, 1987), yet the rest of the consensus is absent. At the 3' end of the intron, a consensus splice acceptor site does occur (14 out of 15 bases identical), yet it is not at the position expected from the cDNA/genomic

sequence match. This probably implies that the cDNA was not derived from the D0485 copy of this gene fragment, but was instead derived from another copy of this sequence, and that the exact position of the splice sites has diverged between the two.

Of the 200bp of the matched region sequenced, there were 33 definite differences between the cDNA and genomic DNA sequences, with a possible 8 others unassigned due to ambiguous sequence data. Of the 33 there were 25 base changes and 8 deleted / inserted bases, including one four base insertion into the genomic DNA sequence. This relatively high rate of changed sequence between the cDNA and genomic DNA suggests that the cDNA was not derived from the D0485 copy of the sequence.

6.8 Open Reading Frames Encoded by the Novel cDNA Products

Additional proof that a DNA sequence is derived from a gene can be acquired by demonstrating that the sequence contains an open reading frame (ORF) which can code for a peptide. The cDNA Family and Product consensus sequences were therefore translated in all three frames and in both directions. ORFs (coding sequences containing no stop or nonsense codons) were obtained for one permutation for Families 3, 4 and 5 and Product 6. These ORFs are shown in Figure 6.15.

Figure 6.15 Open Reading Frames Encoded by Family 3 (A), 4 (B), 5 (C) and Product 6 (D).

A) Family 3 The ORF was produced by translating the reversed consensus sequence, beginning at position 2.

```
1  KDVRNVSART DLAQZQKPQP VNPELTSWCI TGVLDMLNNF RVDSALSTEM  50
51 IPCYISLSED VRYVIFGDDH LSAPTD                          76
```

B) Family 4 Translation of the Family 4 consensus sequence was impossible due to the presence of degenerate bases. Translation of one of its constituents (4-7d) did not yield an ORF; the other (5-7e) also contained degenerate bases. The ORF below resulted from translation of the reversed genomic sequence primed from 488, beginning at position 1. The fact that the cDNA clone sequence did not possess an ORF implies either the existence of sequencing errors in the cDNA clones, or that the cDNA was derived from a transcribed pseudogene, in which mutations had destroyed the ORF.

```
1   KEGWSPKFQD VSFKSVRSSA TVFLPDSSGK EF                    32
```

C) Family 5 This ORF was generated by translating the reversed genomic sequence primed from 491, starting at position 1. As for Family 4, the Family 5 consensus sequence and a constituent clone (2-5c) could not be translated due to the presence of degenerate bases. The other constituent clone, 1-11d, did not possess an ORF.

```
1   LCIDLVSLWH KNKFFWVSSL RKMDLEFPQA FQKELTCLIC LNYLIDPITI   50
51  GCWHSFCRPC                                              60
```

D) Product 6 This ORF was derived by translating the 4-1f cDNA clone sequence, beginning at position 1.

```
1   REPPPPAPGH TALKFLLGLT QCQPPKTSHF PPDISSASRM EVENLGLGPR   50
51  TGDHAGLPWT VTDSGSHQD                                    69
```

The presence or lack of an ORF within a DNA sequence is critically dependent upon that sequence, since only a single wrong base will destroy the continuity. The cDNA consensus sequences and corresponding genomic sequences were commonly derived from a single sequencing reaction, hence errors and ambiguities are inevitable. This could explain why no ORF

was present in the Family 1, 2 or Product 7 sequence. Although the former two have not been categorically proven to be coding sequence, this is not so for Product 7, for which an exon/intron structure has been demonstrated. Unless the Product 7 sequence was derived from a non-processed pseudogene, the lack of ORF must be attributed to sequencing errors.

The necessity for accurate sequence data before confirmation of genuine ORFs may imply that the ORFs mentioned above occur merely a result of chance. None of the ORFs are particularly long, which makes this possibility more realistic. The ORFs from Family 4 and Product 6 were in the orientation that would be predicted, assuming that these cDNA fragments are derived from the same gene as each other and also as Product 7. This is a reasonable assumption since these products map close to each other (within 6.5kb) in genomic DNA and lie adjacent to a putative CpG island (characterised by rare-cutter restriction enzyme sites). The direction of transcription can be surmised by the orientation of the splice sites present in the Product 7 sequence and by the presumed presence of the CpG island at the 5' end of the gene. The ORF from Family 5 was however in the opposite orientation to that expected from the inter-fragment PCRs, and thus the ORF probably occurs by chance.

The degeneracy of the genetic code means that products similar at the amino acid level may appear more diverged at the nucleotide level. Hence a database search between translated DNA fragments and translated database entries can increase the power to find homologies. Such database searches were performed with the seven novel cDNA consensus sequences, translated in all six possible reading frames to minimise the effect of sequencing errors. No significant database homologies were found.

CHAPTER 7

DISCUSSION AND FUTURE PROSPECTS

7.1 Discussion of Results and Related Future Studies

This thesis describes physical mapping of a YAC contig in the region of a translocation breakpoint associated with schizophrenia, a cDNA selection experiment designed to isolate genes from the region around the breakpoint, and then characterisation of the derived cDNAs, to a greater or lesser extent. The physical mapping of the region, including the isolation of end clones, splinkerette PCR and pulsed field mapping of YACs has been discussed in Chapter 3 and therefore will not be alluded to further here.

Since the translocation breakpoint had been shown to be strongly associated with major mental illness in the pedigree (St.Clair *et al*, 1990), it was assumed that the two were causally related, that is that the translocation event was in some way disrupting, altering the expression of or fixing a mutant allele of a gene located either at the breakpoint or nearby. The chromosome 11 side of the translocation breakpoint was chosen for initial investigation, due to previous associations of cytogenetic rearrangements or candidate genes from this chromosome with psychiatric illness (St.Clair *et al*, 1994). Physical mapping in the region of the chromosome 11 translocation breakpoint led to the construction of a 3Mb YAC contig, in which YAC D0485 was proven to span the breakpoint (Evans *et al*, 1995 and references therein).

The YAC D0485 was chosen as the best candidate genomic substrate for a gene-finding experiment. YAC D0485 contains approximately 800kb of DNA proximal and 500kb distal to the translocation breakpoint, which should be a sufficiently large range to detect the gene concerned, even if it was located at some distance from the translocation breakpoint. The cDNA resource

chosen was derived from foetal brain. This was partly because candidate genes for involvement in schizophrenia would most likely be expressed in the developing brain. Also that it is estimated that at least 30% of all genes are expressed in the brain (Sutcliffe, 1988), thus increasing the probability that all relevant genes located within D0485 would be detected by this method.

cDNA selection (HF-CSC) was the gene-finding method of choice, in concert with a complementary EL-CSC experiment and co-association of putative transcripts with CpG island-associated restriction enzyme sites on the chromosome. cDNA selection and EL-CSC are based upon the same principle and the same DNA input resources were employed. The parallel analysis of the overlapping spectrum of products thus generated was useful in ensuring a complete survey of brain-expressed genes in the region. Products unique to both libraries were identified - Family 1 clones were derived solely from the EL library whereas Products 6 and 7 were derived from the HF library and gave no positive signals when hybridised onto the EL library. EL is a more stringent technique than HF, thus more competent in the recovery of rare cDNA species, but its requirement for intra-exonic fragments limits the non-recovery of some genuinely coincident fragments. Since full length cDNAs have not yet been isolated from the products from either library it is impossible to tell whether both libraries do in fact contain different fragments of the same genes.

The most significant drawback of the use of cDNA selection methodology alone was the restriction to isolation of genes expressed in the tissue chosen, even though the problem was minimised by the use of brain cDNA. Alternative strategies such as exon trapping, direct screening of cDNA libraries and extensive sequence analysis are being evaluated for future gene-finding experiments in the region of the two breakpoints. A survey of recent literature reveals that cDNA selection techniques are however frequently the preferred option for finding genes from cloned resources such

as YACs. For instance, cDNA selection was recently utilised in the identification of approximately 50 new genes from chromosome 21 (Tassone *et al*, 1995). Also, during the search for genes from the BRCA1 region on chromosome 17q, cDNA selection was found to be more efficient than direct library screening or genomic sequencing, by contributing to the isolation of 39 of the 45 candidate gene fragments (Harshman *et al*, 1995).

Analysis of the HF and EL libraries led to the identification of a new member of the α-tubulin family, plus several novel cDNA fragments. The exact number of novel genes represented by these clones is not known, but mapping data is consistent with at least two and possibly as many as four. The α-tubulin-related locus maps closest to the translocation breakpoint, approximately 250-300kb away. It appears not to be associated with a CpG island, whereas all the other products map to the region of the YAC which contains three *Not I* sites. Since 89% of *Not I* sites occur within a CpG island (Lindsay and Bird, 1987) this is consistent with the presence of genes in this region. A set of 4 of the novel cDNA fragments map to within 6.5kb of each other and within 10kb of a *Not I* site, a *BssH II* site and two *Eag I* sites, which is highly suggestive that these fragments are derived from the same gene and that that gene is associated with a CpG island.

Since there were no genes already known to map to the D0485 region, it was not possible to evaluate the efficiency of the gene finding experiment with respect to the isolation of known genes as a positive control. The paucity of identified genes implies that this is a gene-poor region of the genome. The average density of human genes is five every 100kb, and hence a total of 65 genes might be expected to map to the 1.3Mb of D0485. The map position of the novel gene fragments is consistent with the distribution of *Not I* sites (all three clustered at the proximal end) and also the position of the YAC on the boundary between a G band and an R band. The almost complete coverage of the α-tubulin gene with cDNA fragment clones

could perhaps be taken as an indicator that the cDNA selection had indeed been very efficient at recovering genuinely coincident cDNA fragments. Similarly the ability to connect randomly picked clones into a set of four by genomic PCR attests to a high efficiency of recovery of cDNA fragments.

The α-tubulin gene was present at a very high frequency in the HF product library, probably greater than 36%. This would indicate either numerous copies of the genomic α-tubulin sequence or a highly expressed cDNA species, or both. Mapping data from D4085 are consistent with only one copy of the α-tubulin sequence within the YAC, and as such the high representation of α-tubulin clones in the library is probably due to abundant α-tubulin cDNA species. All the known α-tubulin genes are sufficiently conserved in the coding sequence that one genomic copy can hybridise strongly to cDNAs derived from other family members; this was reflected in the distribution of foetal brain, keratinocyte and novel α-tubulin cDNAs in the HF product library. Sequencing of the α-tubulin gene from D0485 itself was therefore necessary in order to ascertain which, if any, of the CSC clone sequences was representative of the α-tubulin gene on the YAC. This was achieved through the sequencing of plasmid subclones of α-tubulin positive D0485-derived cosmids, and from the sequencing of PCR products from within the gene. The allele-specificity of the latter was provided by primary amplification from primers outside the gene sequence itself.

Comparison of the D0485 α-tubulin gene with known α-tubulin gene sequences in the database revealed that it was most similar to the keratinocyte α-tubulin gene. Apart from single base changes relative to the keratinocyte sequence, and short deletions and insertions (including one 7 base pair deletion in the coding sequence), the major feature of the D0485 α-tubulin gene was a lack of introns. This, and the fact that its expression could not be detected in a variety of human tissues by allele-specific RT-PCR, suggest that the D0485 α-tubulin gene is, in fact, a processed

pseudogene. The gene sequence itself is capable of coding for a protein very similar to that derived from the keratinocyte α-tubulin gene, although it would be truncated to 80% of its normal length as a result of the change in reading frame after the seven base pair deletion. It is still formally possible therefore that this gene is expressed, but in a highly specific manner and not in the tissues tested.

It is also possible that expression of the D0485 α-tubulin gene is activated or repressed by the presence of the translocation. Ways in which this could be tested and reasons to suppose that α-tubulin could be involved in the aetiology of schizophrenia have already been discussed in Chapter 5.

Three products were probably artefacts of the product library since they did not hybridise back to D0485. All the other fragments recovered and tested did map back to D0485, but none gave matches or significant similarities in the current sequence databases. Sequence analysis allowed them to be grouped into 5 families, leaving two individual products. Two of the families and the two individual products were tentatively grouped together, on the basis of similar hybridisation patterns on mouse and human cDNA and extremely close genomic locations as revealed by the detailed mapping of a cosmid. This group of four products is probably associated with a CpG island. The other three products were linked by long PCR in genomic DNA, but no cosmid was common to all three.

Sequencing of the genomic and cDNA copies of these fragments demonstrated an exon/intron structure for two of the products. This is accepted to be categorical proof that the sequence of interest is derived from a gene. In most cases the cDNA sequence was significantly different from the exon genomic sequence (that is, more differences than can be attributed to sequencing errors). This is consistent with the observation from hybridisation to total genomic DNA revealed that most of the products were

present at other loci in the genome, usually in multiple copies relative to the single D0485 locus.

Preliminary experiments to assess the expression pattern of these cDNA fragments will now be undertaken, by Northern blot and RT-PCR analysis. RNA *in situ* analysis of some of the fragments is also underway, which should reveal their expression pattern in mouse embryos of various gestational ages, and possibly provide an indication of their function.

RT-PCR from a permissive tissue using primers derived from the cDNA fragments could be used to ascertain the total number of genes represented by the novel cDNA fragments. Sequencing of these products would be a direct route to obtaining more coding sequence. Alternatively some further information could be obtained by extending the length of the known cDNA by 5' or 3' rapid amplification of cDNA ends (RACE) (Frohman *et al*, 1988 and Loh *et al*, 1989) and then sequencing the product. Probably the most efficient way to obtain all the information possible from the coding sequence of a gene is to screen a full length cDNA library, and then sequence a positive clone. For this purpose, a full length foetal brain cDNA library has been obtained from the HGMP resource centre. The sequence could then be used to investigate the possible involvement of the genes in schizophrenia and other psychiatric illnesses by screening for mutations in patients.

Assuming that the CSC novel cDNA fragments are proven to correspond to *bona fide* genes by extension of their sequence, studies will be undertaken to determine the function of these genes. This may be partly possible by assessing homology with known genes at the nucleotide or protein level, and by analysis of the protein's hydrophobicity profile and domains indicative of perhaps membrane spanning or DNA binding regions. Use could possibly be made of the syntenic genomic regions in other species, where similar genes may be of known function and/or associated with a phenotype.

Analysis of the encoded proteins themselves may reveal more information about function. Western blotting can be used to determine the size of the protein in different tissues, and the size compared with that expected from analysis of the coding sequence. Immunofluorescence can reveal the localisation of a protein within a cell, and co-immunoprecipitation the interaction with other proteins. Protein-protein interactions can also be detected by the yeast two hybrid system (Fields and Song, 1989).

To gain more knowledge of the function of a gene, an animal model for over or under expression of the gene may be created. Insertion of or replacement with mutated version(s) of the gene at one or both chromosomal copies can reveal phenotypes caused by haploinsufficiency or complete absence of a functional protein. Alternatively transgenic mice could be made containing multiple copies of the normal or mutant human gene, perhaps on a YAC. The presence or absence of an abnormal phenotype in these situations will reveal whether the gene acts in a dominant, recessive or dominant negative fashion and the effect of different mutations (point mutations, gross deletions etcetera) can be tested. Insertion of a reporter gene (for instance β-galactosidase) as a fusion construct into the mouse gene or a human or mouse YAC transgene could be used to follow expression at the single cell level throughout development. Study of normal and abnormal gene expression patterns and physiological study of abnormal phenotypes can lead to an understanding of the function of the gene in the experimental animal.

7.2 Concurrent Mapping of the t(1;11) Translocation Breakpoint and Flanking Genomic Regions

During the course of work for this thesis, other members of the Psychiatric Genetics group have proceeded with the physical mapping and detailed analysis of the translocation breakpoints on chromosomes 1 and 11. On chromosome 11, the cosmids subcloned from YAC D0485 were used to construct a contig across the translocation breakpoint. FISH analysis on metaphase chromosomes derived from a cell line bearing the der(1) and der(11) translocation chromosomes ('MAFLI') was used to prove that individual cosmids spanned the breakpoint on chromosome 11. These were then subcloned further, and a detailed restriction map of the breakpoint region constructed from them. In this way, the region containing the breakpoint was narrowed to a 2.15kb *EcoRI* & *Hind III* fragment. This was proven to contain the breakpoint by FISH analysis on MAFLI and by hybridisation to the somatic cell hybrid panel. This 2.15kb fragment was then sequenced. A trinucleotide repeat within this sequence gave a significant match with an STS mapped to chromosome 1. Although the STS has not been mapped in more detail, it is possible that homology between the breakpoint regions of chromosomes 1 and 11 provides the mechanism for the translocation event. The exact position of the breakpoint was further narrowed, to within 220bp, by PCR amplification of fragments within the 2.15kb from the somatic cell hybrids containing either the der(1) or the der(11) chromosomes.

Splinkerette and Vectorette PCR were used as complementary methods to 'walk' across the breakpoint on the der(1) chromosome. For splinkerette PCR, the somatic cell hybrid containing the der(1) chromosome, digested with *Bgl II* and *Bcl I*, was used as the template for splinkerette ligation and PCR. The sequence specific primer was derived from the sequence of the 2.15kb breakpoint subclone. A 2.7kb product was derived from the *Bcl I* digested DNA, which is the size expected from restriction enzyme analysis of

the normal and translocation chromosomes. This will now be sequenced to locate the exact position of the breakpoint. No attempt has yet been made to walk across the breakpoint on the der(11) chromosome.

Progress has also been made on the mapping of the chromosome 1 breakpoint, largely by a collaborating group in France (led by Benoît Arveiler). Several microdissection clones mapped to chromosome 1 were used to screen YAC libraries, and a YAC contig which spanned the breakpoint was assembled by marker analysis with microdissection clones, YAC end clones and published AFM markers. YACs have been proven to span the translocation breakpoint by FISH. Cosmids across the breakpoint are now being isolated by the construction of a cosmid library from one of the breakpoint-spanning YACs and by screening an ICRF chromosome 1-specific cosmid library (obtained from H. Lehrach) with markers in the immediate vicinity of the breakpoint. It is hoped that independent physical mapping and sequencing of the chromosome 1 breakpoint, plus sequence data gleaned from the breakpoint-spanning splinkerette PCR products will facilitate an understanding of the effects and possibly the mechanism of the translocation.

7.3 How near are we to finding a 'schizophrenia gene'?

The detailed physical mapping of both the translocation breakpoints is almost complete, which will permit exhaustive sequence analysis in the immediate vicinity of the breakpoint, and identification, or not, of a gene whose coding sequence is disrupted by the breakpoint. In this, the simplest case, this gene may be causally associated with mental illness, or the predisposition to it, in this pedigree.

The DNA sequence already determined from the immediate chromosome 11 breakpoint region, plus the absence of CSC products and CpG islands, suggests that there is not a disrupted gene on this side of the breakpoint. No

sequence data is as yet available for the chromosome 1 side of the breakpoint. A single *Not I* site has been detected in the 750kb YAC spanning the chromosome 1 breakpoint, but its exact distance from the breakpoint is not yet known.

If there is no gene disrupted by the translocation breakpoint on either chromosome 1 or 11, then a position effect on a more distant gene is implicated. The cDNA selection search for genes within YAC D0485 has yielded possible candidates. The nearest gene to the breakpoint, α-tubulin, has the characteristics of a pseudogene, and despite plausible mechanisms by which it may still be involved, the studies required to investigate this are complex and require a great investment of time. Even if the translocation event was causing aberrant expression of this gene in psychiatric patients, then the tissue and temporal expression may be so specific as to prohibit detection. The other genes identified on D0485 are novel and therefore require a great deal more study before their potential as candidate genes can be fully assessed, and if necessary discarded.

Further gene finding experiments on chromosome 11 will be carried out, both cDNA selection on YACs further from the breakpoint, and possibly other methods applied to D0485, such as sequence analysis extending outwards from the breakpoint and a computer search for likely coding sequences. Gene searches will also be carried out on the chromosome 1 side of the breakpoint, probably beginning with cDNA selection and direct screening of cDNA libraries with YACs or cosmids close to the breakpoint. Exon trapping may also be utilised as a non expression-based method.

In addition to positional cloning, a candidate gene approach is also being employed in the search for genes. The localisation of two α-actinin genes, ACTN2 and ACTN3, to the same cytogenetic regions as the translocation breakpoints on chromosome 1 and 11 respectively (Beggs *et al*, 1992)

provoked speculation that these genes were affected by the translocation. Probes for ACTN2 and ACTN3 have been amplified by RT-PCR from human foetal limb tissue and their exact locations will be mapped relative to the two translocation breakpoints. For similar reasons as those put forward for α-tubulin, α-actinin can be viewed as a possible candidate for involvement in the aetiology of psychiatric illness. The α-actinins are believed to have diverse cellular functions, probably related to their ability to cross-link actin filaments (Beggs *et al*, 1992). Actin filaments are believed to responsible for the guidance of the growth cone of growing neurons (Sobue, 1993), a disturbance of which may be responsible for the neuropathology seen in some schizophrenics. It is interesting to note that dystrophin, which is a member of the same 'spectrin superfamily' of proteins and is also capable of binding to actin, has previously been implicated in the aetiology of schizophrenia.

The recent description of a gene for Alzheimer's disease on chromosome 1q31-42 (Levy-Lahad *et al*, 1995 a and b, Rogaev *et al*, 1995) provided a second candidate gene in the region. Although the neuropathology seen in Alzheimer's disease is quite distinct from that seen in schizophrenia, the two disorders quite frequently co-occur in the same families. Although the YAC containing this gene has been estimated from linkage maps of chromosome 1 to be between 1 and 3 Mb away from the translocation breakpoint, the possibility of a long range position effect cannot be dismissed. A fragment from the gene will be amplified by RT-PCR from somatic cell hybrids containing the der(1) and der(11) chromosomes to investigate whether expression of the gene is affected by the translocation. Mutation detection will be undertaken to assess whether mutated forms of the gene are associated with presence of the t(1;11) translocation.

A search for genes by several methods moving progressively outward from the breakpoint regions of chromosomes 1 and 11, plus evaluation of

candidate genes as they arise, should lead to the identification of a gene or genes which are involved in the aetiology of mental illness in the t(1;11) pedigree. This however is not an end-point. The relevant gene must be firstly be screened for mutations in other, unrelated schizophrenics. If no mutations are found, as is not unexpected for an aetiologically heterogeneous disorder, then mutation analysis in these other individuals must be performed with structurally or functionally-related genes. Also, it is unlikely that a single gene is responsible in all cases for the complex and varying phenotype of schizophrenia, so an understanding of the influence of modifying or other unrelated genes and the effect of the environment must be sought.

Within the last few years, interest in the field of psychiatric genetics as a whole has grown tremendously. New sophisticated methods of linkage analysis are more adept at detecting genes involved in complex disorders, and the detailed emerging maps of the human genome make positional cloning a simpler task. Progress in these areas is invaluable in resolving the stalemate that had existed in the field ever since the basic principles of the psychiatric genetics were described from family, twin and adoption studies.

Genetic analysis of the dementias, notably Alzheimer's disease, is at the forefront of progress in psychiatric genetics, with mutations in several single genes shown to be causative of the disorder and alleles of the apolipoprotein E gene increasing (ε4) or decreasing (ε2) risk (Selkoe, 1995). This success is partly due to the clear neuropathological features associated with all cases of Alzheimer's disease, enabling objective scoring of the affected or unaffected status. In this respect, robust quantitative or qualitative markers of susceptibility status must be sought for the psychoses, in order that they may ultimately replace the current diagnosis based on the psychological profile alone.

The positional cloning studies around the breakpoints of the balanced translocation described here may yield the description of a gene causally associated with schizophrenia and other major mental illness, albeit in only one family. However a detailed understanding of the multiple and interacting genetic and environmental effects which contribute to the disorder in different people is still placed very firmly in the future.

REFERENCES

Abrams R. and Taylor M.A. (1983) The genetics of schizophrenia: a reassessment using modern criteria. *Am. J. Psychiatry,* **140**, 171-175.

Adams M.D., Kelley J.M., Gocayne J.D., Dubnick M., Polymeropoulos M.H., Xiao H., Merril C.R., Wu A., Olde B., Moreno R.F., Kerlavage A.R., McCombie W.R. and Venter J.C. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science,* **252**, 1651-1656.

Adler L.B., Pachtman E., Franks M., Pecevich N., Waldo C. and Freedman R. (1982) Neurophysiological evidence for a defect in neuronal mechanisms involved in sensory gating in schizophrenia. *Biol. Psychiatry,* **17**, 639-654.

Altschul S.F., Gish W., Miller W., Myers E.M. and Lipman D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Anand R., Riley J.H., Butler R., Smith J.C. and Markham A.F. (1990) A 3.5 genome equivalent multi access YAC library: construction, characterisation and storage. *Nucl. Acids. Res.* **19**, 1951-1956.

Andreasen N.C. and Olsen S. (1982) Negative versus positive schizophrenia. *Arch. Gen. Psychiatry,* **39**, 789-794.

Arévalo M.A., Nieto J.M., Andreu D. and Andreu J.M. (1990) Tubulin assembly probed with antibodies to synthetic peptides. *J. Mol. Biol.*, **214**, 105-129.

Arnold C. and Hodgson I.J. (1991) Vectorette PCR: a novel approach to genomic walking. *PCR Methods and Applications,* **1**, 39-42.

Arnold S.E., Lee V.M., Gur R.E. and Trojanowski J.Q. (1991) Abnormal expression of two microtubule-associated proteins (MAP2 and MAP5) in specific subfields of the hippocampal formation in schizophrenia. *Proc. Natl. Acad. Sci. USA,* **88**, 10850-10854.

Arveiler B. and Porteous D.J. (1991) Amplification of end fragments of YAC recombinants by inverse-polymerase chain reaction. *Technique,* **3**, 24-28.

Aschauer H.N., Aschauer-Treiber G., Isenberg K.E., Todd R.D., Knesvich M.A., Garver D.L., Reich T. and Cloninger C.R. (1990) No evidence for linkage between chromosome 5 markers and schizophrenia. *Hum. Hered.* **40**, 109-115.

Aschauer H.N., Fischer G., Isenberg K.E., Meszaros K., Willinger U., Todd R.D., Beran H., Strobl R., Lang M., Fuchs K., Sieghart W., Reich T. and Cloninger C.R. (1993) No proof of linkage between schizophrenia-related disorders including schizophrenia and chromosome 2q21 region. *Eur. Arch. Psychiatry Clin. Neurosci.* **243**, 193-198.

Asherson P., Mant R., Taylor C., Sargeant M., Collier D., Clements A., Nanko S., Whatley S., Gill M., McGuffin P. and Owen M.(1993) Failure to find linkage between schizophrenia and genetic markers on chromosome 21. *Am. J. Med. Genet.* **48**, 161-165.

Asherson P., Parfitt E., Sargeant M., Tidmarsh S., Buckland P., Taylor C., Clements A., Gill M., McGuffin P. and Owen M. (1992) No evidence for a pseudoautosomal locus for schizophrenia. Linkage analysis of multiply affected families. *Br. J. Psychiatry,* **161**, 63-68.

Asherson P., Walsh C., Williams J., Sargeant M., Taylor C., Clements A., Gill M., Owen M. and McGuffin P. (1994) Imprinting and anticipation. Are they relevant to genetic studies of schizophrenia? *Br. J. Psychiatry,* **164**, 619-624.

Aslanidis C. and de Jong P.J. (1991) Coincidence cloning of Alu-PCR products. *Proc. Natl. Acad .Sci. USA*, **88**, 6765-6769

Audebert S., Koulakoff A., Berwald-Netter Y., Gros F., Denoulet P. and Edde B. (1994) Developmental regulation of polyglutamated a- and b-tubulin in mouse brain neurons. *J. Cell Science,* **107**, 2313-2322.

Axelsson R. and Wahlstrom A. (1984) Chromosome aberrations in patients with paranoid psychosis. *Hereditas,* **100**, 29-31.

Bachneff S.A. (1991) Positron emission tomography and magnetic resonance imaging: a review and a local circuit neurons hypo(dys)function hypothesis of schizophrenia. *Biol. Psychiatry,* **30**, 857-886.

Bailey D.M.D., Carter N.P., de Vos D., Leversha M.A., Perryman M.T. and Ferguson-Smith M.A. (1993) Coincidence painting: a rapid method for cloning region specific DNA sequences. *Nucl. Acids Res.* **21**, 5117-5123.

Baron M. (1976) Albinism and schizophreniform psychosis: a pedigree study. *Am. J. Psychiatry,* **133**, 1070-1073.

Baron M. (1977) Linkage between an X-chromosome marker (deutan color blindness) and bipolar affective illness. Occurrence in the family of a lithium carbonate-responsive schizo-affective proband. *Arch. Gen. Psychiatry,* **34**, 721-725.

Baron M., Endicott J. and Ott J. (1990) Genetic linkage in mental illness. Limitations and prospects. *Br. J. Psychiatry,* **157**, 645-655.

Baron M. and Gruen R.S. (1991) Schizophrenia and affective disorder: are they genetically linked? *Br. J. Psychiatry,* **159**, 267-270.

Barr C.L., Kennedy J.L., Lichter J.B., Van Tol H.H., Wetterberg L., Livak K.J. and Kidd K.K. (1993) Alleles at the dopamine D4 receptor locus do not contribute to the genetic susceptibility to schizophrenia in a large Swedish kindred. *Am. J. Med. Genet.* **48**, 218-222.

Barr C.L., Kennedy J.L., Pakstis A.J., Castiglione C.M., Kidd J.R., Wetterberg L. and Kidd K.K. (1994a) Linkage study of a susceptibility locus for schizophrenia in the pseudoautosomal region. *Schizophr. Bull.* **20**, 277-286.

Barr C.L., Kennedy J.L., Pakstis A.J., Wetterberg L., Sjogren B., Bierut L., Wadelius C., Wahlstrom J., Martinsson T., Giuffra L., Gelernter J., Hallmayer J., Moises H.W., Kurth J., Cavalli-Sforza L.L. and Kidd K.K. (1994b) Progress in a genome scan for linkage in schizophrenia in a large Swedish kindred. *Am. J. Med. Genet.* **54**, 51-58.

Bassett A.S. (1991) Linkage analysis of schizophrenia: challenges and promise. *Social Biology,* **38(3-4)**, 189-196.

Bassett A.S. and Honer W.G. (1994) Evidence for anticipation in schizophrenia. *Am. J. Hum. Genet.* **54**, 864-870.

Bassett A.S., McGillivray B.C., Jones B.D. and Pantzar J.T. (1988) Partial trisomy of chromosome 5 cosegregating with schizophrenia. *Lancet,* **i**, 799-801.

Bebbington P. and Kuipers L. (1994) The clinical utility of expressed emotion in schizophrenia. *Acta Psychiatr. Scand. Suppl.* **382**, 46-53.

Beckwith J.P., Stefanis N.C., McLaughlin D.P. and Kerwin R.W. (1995) The expression of NMDA receptor subunits in schizophrenia post-mortem hippocampus. *Schizophr. Res.* **15(1-2)**, 54

Beggs A.H., Byers T.J., Knoll J.H.M., Boyce F.M., Bruns G.A.P. and Kunkel L.M. (1992) Cloning and characterisation of two human skeletal muscle alpha-actinin genes located on chromosomes one and eleven. *J. Biol. Chem.* **267**, 9281-9288.

Beggs A.H., Hoffman E.P., Snyder J.R., Arahata K., Specht L., Shapiro F., Angelini C., Sugita H. and Kunkel L.M. (1991) Exploring the molecular basis for variability among patients with Becker muscular dystrophy: dystrophin gene and protein studies. *Am. J. Hum. Genet.* **49**, 54-67.

Beratis S., Gabriel J. and Hoidas S. (1994) Age at onset in subtypes of schizophrenic disorders. *Schizophr. Bull.* **20**, 287-296.

Berry R., Stevens T.J., Walter N.A.R., Wilcox A.S., Rubano T., Hopkins J.A., Weber J., Goold R., Soares M.B. and Sikela J.M. (1995) Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. *Nat. Genet.,* **10**, 415-423.

Bird A.P. (1987) CpG islands as gene markers in the vertebrate nucleus. *Trends. Genet.* **3(12)**, 342-347.

Bleich A., Brown S., Kahn R. and van Praag H.M. (1988) The role of serotonin

in schizophrenia. *Schizophr. Bull.* **14(2)**, 297-315.

Blennow G. and McNeil T.F. (1991) Neurological deviations in newborns at psychiatric high risk. *Acta Psychiatr. Scand.* **84**, 179-184.

Boguski M.S., Lowe T.M. and Tolstoshev C.M. (1993) dbEST--database for "expressed sequence tags". *Nat. Genet.* **4**, 332-333.

Boguski M.S. and Schuler G.D. (1995) ESTablishing a human transcript map. *Nat. Genet.* **10**, 369-371.

Bonaldo M.F. (1992) Tubulin alpha-like 2 (TUBA2).*SCW 13.*

Boucher D., Larcher J-C., Gros F. and Denoulet P. (1994) Polyglutamylation of tubulin as a progressive regulator of *in vitro* interactions between the microtubule-associated protein tau and tubulin. *Biochemistry*, **33**, 12471-12477.

Bracha H.S., Torrey E.F., Bigelow L.B., Lohr J.B. and Linington B.B. (1991) Subtle signs of prenatal maldevelopment of the hand ectoderm in schizophrenia: a preliminary monozygotic twin study. *Biol. Psychiatry,* **30**, 719-725.

Brookes A.J. (1994) Identifying and directly purifying transcribed elements. In Hochgeschwender U. and Gardiner K. (eds.), *Identification of Transcribed Sequences.* Plenum Press, New York, pp. 111-121.

Brookes A.J., Slorach E.M., Morrison K.E., Qureshi S.J., Blake D., Davies K. and Porteous D.J. (1994) Cloning the shared components of complex DNA resources. *Hum. Mol. Gen.* **3**, 2011-2017.

Brown P. (1995) Rescuing minds from disease and decay. *New Scientist, 14th Nov (suppl.),* 2-8.

Buckler A.J., Chang D.D., Graw S.L., Brook J.D., Haber D.A., Sharp P.A. and Housman D.E. (1991) Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA,* **88**, 4005-4009.

Burke D.T., Carle G.F. and Olson M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science,* **236**, 806-812.

Butler P.D., Susser E.S., Brown A.S., Kaufman C.A. and Gorman J.M. (1994) Prenatal nutritional deprivation as a risk factor in schizophrenia: preclinical evidence. *Neuropsychopharmacol.* **11(4)**, 227-235.

Cambray-Deakin M.A. and Burgoyne R.D. (1987) Post-translational modifications of a-tubulin:acetylated and detyrosinated forms in axons of rat cerebellum. *J. Cell Biology,* **104**, 1569-1574.

Campion D., d'Amato T., Bastard C., Laurent C., Guedj F., Jay M., Dollfus S., Thibaut F., Petit M., Gorwood P., Babron M.C., Waksman G., Martinez M. and Mallet J. (1994) Genetic study of dopamine D1, D2, and D4 receptors in schizophrenia. *Psychiatry Res.* **51**, 215-230.

Cannon T.D. (1991) Genetics and perinatal sources and structural brain abnormalities in schizophrenia. In Mednick S.A., Cannon T.D., Barr C.E. and Lyon M. (eds.), *Fetal neural development and adult schizophrenia.* Cambridge University Press, New York, pp. 174-178.

Cannon T.D. and Marco E. (1994) Structural brain abnormalities as indicators of vulnerability to schizophrenia. *Schizophr. Bull.* **20**, 89-102.

Cantor-Graae E., McNeil T.F., Sjostrom K., Nordstrom L.G. and Rosenlund T. (1994) Obstetric complications and their relationship to other etiological risk factors in schizophrenia. A case-control study. *J. Nerv. Ment. Dis.* **182**, 645-650.

Carter M. and Watts C.A.H. (1971) Possible biological advantages among schizophrenics' relatives. *Br. J. Psychiatry,* **118**, 453-460.

Castle D.J. and Murray R.M. (1991) The neurodevelopmental basis of sex differences in schizophrenia. *Psychol. Med.* **21**, 565-575.

Chou Q., Russell M., Birch D.E., Raymond J. and Bloch W. (1992) Prevention of PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucl. Acids Res.* **20**, 1717-1723.

Chumakov I., Rigault P., Guillou S., Ougen P., Billaut A., Guasconi G., Gervy P., LeGall I., Soularue P., Grinas L., Bougueleret L., Bellané-Chantelot C., Lacroix B., Barillot E., Gesnouin P., Pook S., Vaysseix G., Frelat G., Schmitz A., Sambucy J-L., Bosch A., Estivil X., Weissenbach J., Vignal A., Riethman H., Cox D., Patterson D., Gardiner K., Hattori M., Sakaki Y., Ichikawa H., Ohki M., Le Paslier D., Heilig R., Antonarakis S. and Cohen D. (1992) Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature,* **359**, 380-387.

Clarke D.J. and Buckley M.E. (1989) Familial association of albinism and schizophrenia. *Br. J. Psychiatry,* **155**, 551-553.

Claverie J. (1994) A streamlined random sequencing strategy for finding coding exons. *Genomics,* **23**, 575-581.

Cleghorn J.M., Zipursky R.B. and List S.J. (1991) Structural and functional brain imaging in schizophrenia. *J. Psychiatry Neurosci.* **16**, 53-74.

Cleveland D.W. and Sullivan K.F. (1985) Molecular biology and genetics of tubulin. *Annu. Rev. Biochem.* **54**, 331-365.

Cloninger C.R. (1994) Turning point in the design of linkage studies of schizophrenia. *Am. J. Med. Genet.* **54**, 83-92.

Cohen D., Chumakov I. and Weissenbach J. (1993) A first-generation physical map of the human genome. *Nature,* **366**, 698-701.

Collinge J., DeLisi L.E., Boccio A., Johnstone E.C., Lane A., Larkin C., Leach M., Lofthouse R., Owen F., Poulter M., Shah T., Walsh C. and Crow T.J. (1991) Evidence for a pseudo-autosomal locus for schizophrenia using the method of affected sib pairs. *Br. J. Psychiatry,* **158**, 624-629.

Collins F.S. (1992) Positional cloning: let's not call it reverse anymore [news]. *Nat. Genet.* **1**, 3-6.

Collins J. and Hohn B. (1978) Cosmids: a type of plasmid gene cloning vector that is packageable in vitro in bacteriophage lambda heads. *Proc. Natl. Acad. Sci. USA,* **75**, 4242-4246.

Coon H., Hoff M., Holik J., DeLisi L., Crowe T., Freedman R., Shields G., Boccio A.M., Lerman M., Gershon E.S., Gejman P.V., Leppert M. and Byerly W. (1993)

C to T nucleotide substitution on codon 713 of amyloid precursor protein gene not found in 86 unrelated schizophrenics from multiplex families. *Am. J. Med. Genet.* **48**, 36-39.

Coon H., Holik J., Hoff M., Reimherr F., Wender P., Myles Worsley M., Waldo M., Freedman R. and Byerley W. (1994) Analysis of chromosome 22 markers in nine schizophrenia pedigrees. *Am. J. Med. Genet.* **54**, 72-79.

Coon H., Jensen S., Hoff M., Holik J., Plaetke R., Reimherr F., Wender P., Leppert M. and Byerley W. (1993) A genome-wide search for genes predisposing to manic-depression, assuming autosomal dominant inheritance. *Am. J. Hum. Genet.* **52**, 1234-1249.

Coon H., Jensen S., Holik J., Hoff M., Myles Worsley M., Reimherr F., Wender P., Waldo M., Freedman R., Leppert M. and Byerley W. (1994) Genomic scan for genes predisposing to schizophrenia. *Am. J. Med. Genet.* **54**, 59-71.

Coon H., Sobell J., Heston L., Sommer S., Hoff M., Holik J., Umar F., Robertson M., Reimherr F., Wender P., Vest K., Myles-Worsley M., Gershon E.S., DeLisi L.E., Shields G., Dale P.W., Polloi A., Waldo M., Leonard S., Sikela J., Freedman R. and Byerley W. (1994) Search for mutations in the beta 1 GABAA receptor subunit gene in patients with schizophrenia. *Am. J. Med. Genet.* **54**, 12-20.

Cowan N.J., Dobner P.R., Fuchs E.V. and Cleveland D.W. (1983) Expression of human a-tubulin genes: interspecies conservation of 3' untranslated regions. *Mol. Cell. Biol.* **3**, 1738-1745.

Crampton J.M., Davies K.E. and Knapp T.F. (1981) The occurrence of families

of repetitive sequences in a library of cloned cDNA fom human lymphocytes. *Nucl. Acids Res.* **9**, 3821-3834.

Crocq M.A., Mant R., Asherson P., Williams J., Hode Y., Mayerova A., Collier D., Lannfelt L., Sokoloff P., Gill M., Macher J.P., McGuffin P. and Owen M.J. (1992) Association between schizophrenia and homozygosity at the dopamine D3 receptor gene. *J. Med. Genet.* **29**, 858-860.

Cross S.H., Charlton J.A., Nan X. and Bird A.P. (1994) Purification of CpG islands using a methylated DNA binding column. *Nat. Genet.* **6**, 236-244.

Crow T.J. (1980a) Molecular pathology of schizophrenia: more than one disease process. *BMJ,* **280**, 66-68.

Crow T.J. (1980b) Temporal lobe assymetries as the key to the etiology of schizophrenia. *Schizophr. Bull.* **16(3)**, 433-443.

Crow T.J., Ball J., Bloom S.R., Brown R., Bruton C.J., Colter N., Frith C.D., Johnstone E.C., Owens D.G.C. and Roberts G.W. (1989) Schizophrenia as an anomaly of development of cerebral assymetry. *Arch. Gen. Psychiatry,* **46**, 1145-1150.

Crow T.J., DeLisi L.E., Lofthouse R., Poulter M., Lehner T., Bass N., Shah T., Walsh C., Boccio Smith A., Shields G. and Ott J. (1994) An examination of linkage of schizophrenia and schizoaffective disorder to the pseudoautosomal region (Xp22.3). *Br. J. Psychiatry,* **164**, 159-164.

Crow T.J. and Done D.J. (1992) Prenatal exposure to influenza does not cause schizophrenia. *Br. J. Psychiatry,* **161**, 390-393.

Crow T.J., Poulter M., Lofthouse R., Chen G., Shah T., Bass N., Morganti C., Vita A., Smith C., Boccio Smith A., Shields G. and DeLisi L.E. (1993) Male siblings with schizophrenia share alleles at the androgen receptor above chance expectation. *Am. J. Med. Genet.* **48**, 159-160.

Crowe R.R., Black D.W., Wesner R., Andreasen N.C., Cookman A. and Roby J. (1991) Lack of linkage to chromosome 5q11-q13 markers in six schizophrenia pedigrees. *Arch. Gen. Psychiatry,* **48**, 357-361.

Cuesta M.J. and Peralta V. (1994) Lack of insight in schizophrenia. *Schizophr. Bull.* **20(2)**, 359-366.

d'Amato T., Campion D., Gorwood P., Jay M., Sabate O., Petit C., Abbar M., Malafosse A., Leboyer M., Hillaire D., Clerget-Darpoux F., Feingold J., Waksman G. and Mallet J. (1992) Evidence for a pseudoautosomal locus for schizophrenia. II: Replication of a non-random segregation of alleles at the DXYS14 locus. *Br. J. Psychiatry,* **161**, 59-62.

D'Arcangelo G., Miao G.G., Chen S.C., Soares H.D., Morgan J.I. and Curran T. (1995) A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. *Nature,* **374**, 719-723.

Dausset J., Ougen P., Abderrahim H., Billault A., Sambucy J.L., Cohen D. and Le Paslier D. (1992) The CEPH YAC library. *Behring. Inst. Mitt.* 13-20.

Davies T. (1994) Psychosocial factors and relapse of schizophrenia. *BMJ.* **309**, 353-354.

Davis K.L. (1991) Dopamine in schizophrenia: a review and reconceptualization. *Am. J. Psychiatry,* **148**, 1474-1486.

Davison M.D., Baron M.D., Critchley D.R. and Wootton J.C. (1989) Structural analysis of homologous repeated domains in alpha-actinin and spectrin. *Int. J. Biol. Macromol.* **11**, 81-90.

Dawson E.B., Moore T.D. and McGanity W.J. (1970) The mathematical relationship of drinking water lithium and rainfall to mental hospital admission. *Diseases of the Nervous System,* **12**, 811-820.

deLeon J., Dadvand M., Canuso C., White A.O., Stanilla J.K. and Simpson G.M. (1995) Schizophrenia and smoking: an epidemiological survey in a state hospital. *Am. J. Psychiatry,* **152(3)**, 453-455.

DeLisi L.E. (1992) The significance of age of onset for schizophrenia. *Schizophr. Bull.* **18**, 209-215.

DeLisi L.E., Devoto M., Lofthouse R., Poulter M., Smith A., Shields G., Bass N., Chen G., Vita A., Morganti C., Ott J. and Crow T.J. (1994) Search for linkage to schizophrenia on the X and Y chromosomes. *Am. J. Med. Genet.* **54**, 113-121.

Detera-Wadleigh S.D., Berettini W.H., Goldin L.R., Boorman D., Anderson S. and Gershon E. (1987) Close linkage of the c-Harvey-ras-1 and the insulin gene to affective disorder is ruled out in three North American pedigrees. *Nature,* **325**, 806-808.

Detera-Wadleigh S.D., Goldin S.R., Sherrington R., Encio I., Miguel C., de Berrettini W., Gurling H. and Gershon E.S. (1989) Exclusion of linkage to 5q11-

13 in families with schizophrenia and other psychiatric disorders. *Nature,* **340**, 391-393.

Devereux J., Haeberli P. and Smithies O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387-395.

Devon R.S. and Brookes A.J. (1996) Coincidence cloning: taking the coincidences out of genome analysis. *Molecular Biotechnology*, in press.

Devon R.S., Porteous D.J. and Brookes A.J. (1995) Splinkerettes - improved vectorettes for greater efficiency in PCR walking. *Nucl. Acids Res.* **23**, 1644-1645.

Dominguez O. and Lopez-Larrea C. (1994) Gene walking by unpredictably primed PCR. *Nucl. Acids Res.* **22**, 3247-3248.

Don R.H., Cox P.T., Wainwright B.J., Baker K. and Mattick J.S. (1991) "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucl. Acids Res.* **19**, 4008

Donaldson J. (1987) The physiopathologic significance of manganese in the brain: its relation to schizophrenia and neurodegenerative disorders. *Neurotoxicology,* **8**, 457-462.

Done D.J., Crow T.J., Johnstone E.C. and Sacker A. (1994) Childhood antecedents of schizophrenia and affective illness: social adjustment at ages 7 and 11. *BMJ.* **309**, 699-703.

Douglass A.B., Shipley J.E., Haines R.F., Scholten R.C., Dudley E. and Tapp A.

(1993) Schizophrenia, narcolepsy, and HLA-DR15, DQ6. *Biol. Psychiatry,* **34**, 773-780.

Drmanac R., Petrovic N., Glisin V. and Crkvenjakov R. (1986) A calculation of fragments lengths obtainable from human DNA with 78 restriction enzymes: an aid for cloning and mapping. *Nucl. Acids Res.* **14**, 4691-4692.

Duyk G.M., Kim S., Myers R.M. and Cox D.R. (1990) Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl. Acad. Sci. USA,* **87**, 8995-8999.

Eaton W.W. (1991) Update on the epidemiology of schizophrenia. *Epidemiol. Rev.* **13**, 320-328.

Egeland J.A., Gerhard D.S., Pauls D.L., Sussex J.N., Kidd K.K., Allen C.R., Hostetter A.M. and Housman D.E. (1987) Bipolar affective disorders linked to DNA markers on chromosome 11. *Nature,* **325**, 783-787.

Elvin P., Slynn G., Black D., Graham A., Butler R., Riley J., Anand R. and Markham A.F. (1990) Isolation of cDNA clones using yeast artificial chromosome probes. *Nucl. Acids Res.* **18**, 3913-3917.

Erlenmeyer-Kimling L. and Paradowski W. (1966) Selection and Schizophrenia. *The American Naturalist,* **100(916)**, 651-665.

Erlenmeyer-Kimling L., Rainer J.D. and Kallman F.J. (1966) Current reproductive trends in schizophrenia. In Hoch P. and Zubin J. (eds.), *Psychopathology of schizophrenia.* Grune and Stratton, New York, pp. 252-276.

Evans G.A., Lewis K. and Rothenburg B.E. (1989) High efficiency vectors for cosmid microcloning and genomic analysis. *Gene,* **79**, 9-20.

Evans K.L., Brown J., Shibasaki Y., Devon R.S., He L., Arveiler B., Christie S., Maule J.C., Baillie D., Slorach E.M., Anderson S.M., Gosden J.R., Petit J., Weith A., Gosden C.M., Blackwood D.H.R., St.Clair D.M., Muir W.M., Brookes A.J. and Porteous D.J. (1995) A Contiguous Clone Map over 3Mb on the Long Arm of Chromosome 11 across a Balanced Translocation Associated with Schizophrenia. *Genomics,* **28(3)**, 420-428.

Fantes J., Redeker B., Breen M., Boyle S., Brown J., Fletcher J., Jones S., Bickmore W., Fukushima Y., Mannens M., Danes S., van Heyningen V. and Hanson I. (1995) Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Hum. Mol. Gen.* **4(3)**, 415-422.

Faraone S.V. and Tsuang M.T. (1985) Quantitative models of the genetic transmission of schizophrenia. *Psychological Bulletin,* **98(1)**, 41-66.

Feinberg A.P. and Vogelstein A. (1983) A technique for radiolabelling DNA restriction fragments to high specific activity. *Anal. Biochem.* **132**, 6-13.

Feinberg A.P. and Vogelstein A. (1984) A technique for radiolabelling DNA fragments to high specific activity. *Anal. Biochem.* **137**, 266-267.

Fields S. and Song O. (1989) A novel genetic system to detect protein-protein interactions. *Nature,* **340**, 245-246.

Fischer M. (1971) Psychosis in the offspring of schizophrenia monozygotic twins and their normal co-twins. *Br. J. Psychiatry,* **118**, 43-52.

Fischer M., Harvald B. and Hauge M. (1969) A Danish twin study of schizophrenia. *Br. J. Psychiatry,* **115**, 981-990.

Fletcher J.M., Evans K.L., Baillie D., Byrd P., Hanratty D., Leach S., Julier C., Gosden J.R., Muir W., Porteous D.J., St.Clair D. and van Heyningen V. (1993) Schizophrenia-associated chromosome 11q21 translocation: identification of flanking markers and development of chromosome 11q fragment hybrids as cloning and mapping resources. *Am. J. Hum. Genet.* **52**, 478-490.

Foster H.D. (1988) The geography of schizophrenia: possible links with selenium and calcium deficiencies, inadequate exposure to sunlight and industrialisation. *J. Orthomolecular Medicine,* **3**, 135-140.

Frankenburg F.R. (1994) History of the development of antipsychotic medication. *Psych. Clin. North Am.* **17(3)**, 531-540.

Freeman H. (1994) Schizophrenia and city residence. *Br. J. Psychiatry Suppl.* 39-50.

Frohman M.A., Dush M.K. and Martin G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucletide primer. *Proc. Natl. Acad. Sci. USA,* **85**, 8998-9002.

Fukuda R., Hattori M., Sasaki T., Kazamatsuri H., Kuwata S., Shibata Y. and Nanko S. (1993) No evidence for a point mutation at codon 713 and 717 of amyloid precursor protein gene in Japanese schizophrenics. *Jpn. J. Hum. Genet.* **38**, 407-411.

Fulker D.W., Cherny S.S. and Cardon L.R. (1995) Multipoint interval mapping of quantitative trait loci, using sib pairs. *Am. J. Hum. Genet.* **56**, 1224-1233.

Gallant D.M. (1990) Diagnosis of the schizophrenic disorders. *Psychiatr. Med.* **8**, 21-40.

Gattaz W.F., Hubner C.V., Nevalainen T.J., Thuren T. and Kinnunen P.K. (1990) Increased serum phospholipase A2 activity in schizophrenia: a replication study. *Biol. Psychiatry,* **28**, 495-501.

Gatti R.A., Shaked R., Wei S., Mohandas T.K. and Salser W. (1987) Biallelic DNA polymorphism of an alpha-tubulin gene family member on chromosome 12 (TUBA/MspI/2.2;2.0kb). *Nucl. Acids Res.* **15**, 8119

Genest P., Dumas L. and Genest F.B. (1976) Translocation chromosomique t(2;18)(q21;q23) chez un individu schizophrenie et sa fille. *Union Med. Can.* **105**, 1676-1681.

Gerhard D.S., Dobner P.R. and Bruns G. (1985) Testis specific alpha tubulin is on chromosome 2q. *Cytogenet. Cell. Genet.* **40**, 639-640.

Gershon E.S. and Rieder R.O. (1992) Major disorders of mind and brain. *Sci. Am.* **267**, 126-133.

Ghosh S., Palmer S.M., Rodrigues N.R., Cordell H.J., Hearne C.M., Cornall R.J., Prins J.B., McShane P., Lathrop G.M., Peterson L.B., Wicker L.S. and Todd J.A. (1993) Polygenic control of autoimmune diabetes in nonobese diabetic mice. *Nat. Genet.* **4**, 404-409.

Gill M., McGuffin P., Parfitt E., Mant R., Asherson P., Collier D., Vallada H., Powell J., Shaikh S., Taylor C., Sargeant M., Clements A., Nanko S., Takazawa N., Llewellyn D., Pilowsky L., Brush Y., Williams J., Whatley S., Murray R. and Owen M. (1993) A linkage study of schizophrenia with DNA markers from the long arm of chromosome 11. *Psychol. Med.* **23**, 27-44.

Goate A., Chartier Harlin M.C., Mullan M., Brown J., Crawford F., Fidani L., Giuffra L., Haynes A., Irving N., James L., Mant R., Newton P., Rooke K., Roques P., Talbot C., Pericak-Vance M., Roses A., Williamson R., Rosor M., Owen M. and Hardy J. (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature,* **349**, 704-706.

Goodman A.B. (1994) A family history study of schizophrenia spectrum disorders suggests new candidate genes in schizophrenia and autism. *Psychiatr. Q.* **65**, 287-297.

Goodman A.B. (1995) Chromosomal locations and modes of action of genes of the retinoid (vitamin A) system support their involvement in the etiology of schizophrenia. *Am. J. Med. Gen.* **60**, 335-348.

Gorwood P., Leboyer M., d'Amato T., Jay M., Campion D., Hillaire D., Mallet J. and Feingold J. (1992) Evidence for a pseudoautosomal locus for schizophrenia. I: A replication study using phenotype analysis. *Br. J. Psychiatry,* **161**, 55-58.

Goss S.J. and Harris H. (1975) New method for mapping genes in human chromosomes. *Nature,* **225**, 680-684.

Gottesman I.I. and Shields J. (1982) *Schizophrenia: The Epigenetic Puzzle.* Cambridge University Press, New York.

Grandy D.K., Narchionni M.A., Makam H., Stotko R.E., Alfano M., Frothingham L., Fischer J.B., Burke-Howie K.J., Bunzow J.R., Serrer A.C. and Civelli O. (1989) Cloning of the cDNA and gene for a human D2 dopamine receptor. *Proc. Natl. Acad. Sci. USA,* **86**, 9762-9766.

Gyapay G., Morissette J., Vignal A., Dib C., Fizames C., Millasseau P., Marc S., Bernardi G., Lathrop M. and Weissenbach J. (1994) The 1993-94 Genethon human genetic linkage map. *Nat. Genet.* **7**, 246-339.

Hale T. (1993) Will the new antipsychotics improve the treatment of schizophrenia? *BMJ.* **307**, 749-750.

Hall J.L. and Cowan N.J. (1985) Structural features and restricted expression of a human a-tubulin gene. *Nucl. Acids Res.* **13**, 207-223.

Hallmayer J., Kennedy J.L., Wetterberg L., Sjogren B., Kidd K.K. and Cavalli Sforza L.L. (1992) Exclusion of linkage between the serotonin2 receptor and schizophrenia in a large Swedish kindred. *Arch. Gen. Psychiatry,* **49**, 216-219.

Hanson I.M., Poutska A. and Trowsdale J. (1991) New genes in the class II region of the human major histocompatibility complex. *Genomics,* **10**, 412-424.

Harrison G. and Mason P. (1993) Schizophrenia--falling incidence and better outcome? *Br. J. Psychiatry,* **163**, 535-541.

Harshman K., Bell R., Rosenthal J., Katcher H., Miki Y., Swenson J., Gholami Z., Frye C., Ding W., Dayananth P., Eddington K., Norris F.H., Bristow P.K., Phelps R., Hattier T., Stone S., Shaffer D., Bayer S., Hussey C., Tran T., Richardson K., Dehoff B., Lai M., Rosteck P.R., Skolnick M.H., Shattuck-Eidens D. and Kamb A. (1995) Comparison of the positional cloning methods used to isolate the BRCA1 gene. *Hum. Mol. Gen.* **4**, 1259-1266.

Hearne C.M., Ghosh S. and Todd J.A. (1992) Microsatellites for linkage analysis of genetic traits. *Trends. Genet.* **8**, 288-294.

Hegarty J.D., Baldessarini R.J., Tohen M., Waternaux C. and Oepen G. (1994) One hundred years of schizophrenia: a meta-analysis of the outcome literature. *Am. J. Psychiatry,* **151**, 1409-1416.

Hemmings G. (1990) Causes of schizophrenia. *Nutr. Health,* **7**, 11-19.

Heston L.L. (1966) Psychiatric disorders in foster home reared children of schizophrenic mothers. *Br. J. Psychiatry,* **112**, 819-825.

Hodge S.E. (1994) What association analysis can and cannot tell us about the genetics of complex disease. *Am. J. Med. Genet.* **54**, 318-323.

Hodgkinson S., Sherrington R., Gurling H., Marchbanks R., Reeders S., Mallet J., McInnis M., Petursson H. and Brynjolffson J. (1987) Molecular genetic evidence for heterogeneity in manic depression. *Nature,* **325**, 783-787.

Holden R.J., Mooney P.A. and Newman J.C. (1994) Schizophrenia: an extended etiological explanation. *Med. Hypotheses.* **42**, 115-123.

Holland G. and Gosden C. (1990) A balanced chromosomal translocation partially segregating with psychotic illness in a family. *Psych. Res.* **32**, 1-8.

Holzman P.S. (1992) Behavioral markers of schizophrenia useful for genetic studies. *J. Psychiatr. Res.* **26**, 427-445.

Holzman P.S., Proctor L.R., Levy D.L., Yasillo N.J., Meltzer H.Y. and Hurt S.W. (1974) Eye tracking dysfunction in schizophrenic patients and their relatives. *Arch. Gen. Psychiatry,* **31**, 143-151.

Horrobin D.F., Glen A.I. and Vaddadi K. (1994) The membrane hypothesis of schizophrenia. *Schizophr. Res.* **13**, 195-207.

Horrobin D.F. and Huang Y. (1983) Schizophrenia: the role of abnormal essential fatty acid and prostaglandin metabolism. *Medical Hypotheses,* **10**, 329-336.

Hozier J., Graham R., Westfall T., Siebert P. and Davis L. (1994) Preparative *in situ* hybridisation: selection of chromosome region-specific libraries on mitotic chromosomes. *Genomics,* **19**, 441-447.

Humphries C.R., Virgo L., Mortimer A., Barnes T., Hirsch S. and de Belleroche J. (1995) The expression of the N-methyl-D-aspartate receptor subunit NR-1 is decreased in schizophrenia. *Schizophr. Res.* **15(1-2)**, 60

Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell,* **72**, 971-983.

Huxley J., Mayr E., Osmond H. and Hoffer A. (1964) Schizophrenia as a genetic morphism. *Nature*, **204**, 220-221.

Hyde T.M., Nawroz S., Goldberg T.E., Bigelow L.B., Strong D., Ostrem J.L., Weinberger D.R. and Kleinman J.E. (1994) Is there cognitive decline in schizophrenia? A cross-sectional study. *Br. J. Psychiatry,* **164**, 494-500.

Ioannou P.A., Amemiya C.T., Garnes J., Kroisel P.M., Shizuya H., Chen C., Batzer M.A. and de Jong P.J. (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat. Genet.* **6**, 84-89.

Ishida T., Yoneda H., Sakai T., Nonomura Y., Inayama Y., Kono Y. and Kobayashi S. (1993) Pseudoautosomal region in schizophrenia: sex concordance of the affected sibpairs and the association study with DNA markers. *Am. J. Med. Genet.* **48**, 151-155.

Item C. and Sieghart W. (1994) Binding of γ-aminobutyric $acid_A$ receptors to tubulin. *J. Neurochem.*, **63(3)**, 1119-1125.

Itokawa M., Arinami T., Futamura N., Hamaguchi H. and Toru M. (1993) A structural polymorphism of human dopamine D2 receptor, D2(Ser311-->Cys). *Biochem. Biophys. Res. Commun.* **196**, 1369-1375.

James M.R., Richard III C.W., Schott J., Yousry C., Clark K., Bell J., Terwilliger J.D., Hazan J., Dubay C., Vignal A., Agrapart M., Imai T., Nakamura Y., Polymeropoulos M., Weissenbach J., Cox D.R. and Lathrop G.M. (1994) A radiation hybrid map of 506 STS markers spanning human chromosome 11. *Nat. Genet.,* **8**, 70-76.

Jayaram B. and Haley B.E. (1994) Identification of peptides within the base binding domains of the GTP- and ATP-specific binding sites of tubulin. *J. Biol. Chem.*, **269(5)**, 3233-3242.

Johnstone E.C. (1993) Schizophrenia: problems in clinical practice. *Lancet,* **341**, 536-538.

Johnstone E.C., Crow T.J., Frith D.C., Husband J. and Krel L. (1976) Cerebral ventricular size and cognitive impairment in schizophrenia. *Lancet,* **ii**, 924.

Jones C.P., Janson M. and Nordenskjold M. (1989) Separation of yeast chromosomes in the megabase range suitable as size markers for pulsed field gel electrophoresis. *Technique,* **1**, 90-95.

Jones C.T., Morris S., Yates C.M., Moffoot A., Sharpe C., Brock D.J.H. and St.Clair D. (1992) Mutation in codon 713 of the B amyloid precursor protein gene presenting with schizophrenia. *Nat. Genet.* **1**, 306-309.

Jones D.S.C. and Schofield J.P. (1990) A rapid method for isolating high quality plasmid DNA suitable for DNA sequencing. *Nucl. Acids Res.* **18**, 7463-7464.

Jones K.W., Shapero M.H., Chevrette M. and Fournier R.E. (1991) Subtractive hybridization cloning of a tissue-specific extinguisher: TSE1 encodes a regulatory subunit of protein kinase. *Cell,* **66**, 861-872.

Jones P., Rodgers B., Murray R. and Marmot M. (1994) Child development risk factors for adult schizophrenia in the British 1946 birth cohort. *Lancet,* **344**, 1398-1402.

Jordan E. (1992) The Human Genome Project: where did it come from, where is it going? *Am. J. Hum. Genet.* **51**, 1-6.

Joseph R., Dou D. and Tsang W. (1994) Molecular cloning of a novel mRNA (neuronatin) that is highly expressed in neonatal mammalian brain. *Biochem. Biophys,Res. Commun.* **201**, 1227-1234.

Julier C., Hyer R.N., Davies J., Merlin F., Soularue P., Briant L., Cathelineau G., Deschamps I., Rotter J.I., Froguel P., Boitard C., Bell J.I. and Lathrop G.M. (1991) Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. *Nature,* **354**, 155-159.

Kalman M., Kalman E.T. and Cashel M. (1990) Polymerase chain reaction (PCR) with a single primer. *Biochem. Biophys. Res. Commun.* **167**, 504-506.

Kalsi G., Brynjolfsson J., Butler R., Sherrington R., Curtis D., Sigmundsson T., Read T., Murphy P., Sharma T., Petursson H. and Gurling H.M.D. (1995) Linkage analysis of chromosome 22q12-13 in a United Kingdom/Icelandic sample of 23 multiplex schizophrenia families. *Am. J. Med. Gen.* **60**, 298-301.

Keefe R.S., Silverman J.M., Siever L.J. and Cornblatt B.A. (1991) Refining phenotype characterization in genetic linkage studies of schizophrenia. *Soc. Biol.* **38**, 197-218.

Kelsoe J.R., Ginns E.I., Egeland J.A., Gerhard D.S., Goldstein A.M., Bale S.J., Pauls D.L., Long R.T., Kidd K.K., Conte G., Housman D.E. and Paul S.M. (1989) Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. *Nature,* **342**, 238-243.

Kendell R.E. and Kemp I.W. (1989) Maternal influenza in the etiology of schizophrenia. *Arch. Gen. Psychiatry,* **46**, 878-882.

Kendler K.S., Gruenberg A.M. and Tsuang M.T. (1985) Psychiatric illness in first degree relatives of schizophrenic and surgical control patients. *Arch. Gen. Psychiatry,* **42**, 770-779.

Kennedy J.L., Giuffra L.A., Moises H.W., Cavalli-Sforza L.L., Pakstis A.J., Kidd J.R., Casriglione C.M., Sjogren B., Wetterburg L. and Kidd K.K. (1988) Evidence against linkage of schizophrenia to markers on chromosome 5 in a Northern Swedish pedigree. *Nature,* **336**, 167-170.

Kerwin R.W. (1993) Glutamate receptors, microtubule associated proteins and developmental anomaly in schizophrenia: an hypothesis. *Psychol. Med.* **23**, 547-551.

Kerwin R.W. and Harrison P.J. (1995) Loss of non NMDA receptors in medial temporal lobe - a robust neurochemical finding: autoradiographical, gene expression and immunocytochemical findings. *Schizophr. Res.* **15(1-2)**, 62.

Kety S.S., Rosenthal D., Wender P.H., Schulsinger F. and Jacobson B. (1975) Mental illness in the biological and adoptive families of adopted individuals who have become schizophrenic: a preliminary report based on psychiatric interviews. *Proc. Am. Psychopathol. Soc.* **63**, 147-165.

Kety S.S., Rosenthal D., Wender P.H., Schulsinger F. and Jacobson B. (1978) The biologic and adoptive families of individuals who become schizophrenic: prevalance of mental illness and other characteristics. In Wynne L.C., Cromwell

R.L. and Matthysse S. (eds.), *The Nature of Schizophrenia.* John Wiley and Sons Inc. New York, pp. 25-37.

King M., Coker E., Leavey G., Hoare A. and Johnson Sabine E. (1994) Incidence of psychotic illness in London: comparison of ethnic groups. *BMJ.* **309**, 1115-1119.

Knight J.G., Knight A. and Pert C.B. (1987) In Helmchen H. and Henn F.A. (eds.), *Biological perspectives of schizophrenia.* Wiley, New York.

Kohn M.L. (1976) The interaction of social class and other factors in the etiology of schizophrenia. *Am. J. Psychiatry,* **133(2)**, 177-180.

Korn B., Sedlacek S., Manca A., Kioschis P., Konecki P., Lehrach H. and Poutska A. (1992) A strategy for the selection of transcribed sequences in the Xq28 region. *Hum. Mol. Gen.* **1**, 235-242.

Kruglyak L. and Lander E.S. (1995) High-resolution mapping of complex traits. *Am. J. Hum. Genet.*, **56**, 1212-1223.

Kunugi H., Nanko S. and Takei N. (1992) Influenza and schizophrenia in Japan. *Br. J. Psychiatry,* **161**, 274-275.

Kunugi H., Nanko S., Takei N., Saito K., Murray R.M. and Kazamatsuri H. (1995) Small head circumference at birth and schizophrenia. *Schizophr. Res.* **15(1-2)**, 192

Kunugi H., Takei N., Nanko S., Saito K. and Kazamatsuri H. (1994) Head circumference at birth and schizophrenia. *Br. J. Psychiatry,* **165**, 551-559.

Lannfelt L., Sokoloff P., Martres M.P., Pilon C., Giros B., Jonsson E., Sedvall G. and Schwartz J.G. (1992) a a substitution in the dopamine D3 receptor as a useful polymorphism for investigating psychiatric disorders. *Psych. Gen.* **2**, 249-256.

Lapensee M.A. (1992) A review of schizoaffective disorder: I. Current concepts. *Can. J. Psychiatry,* **37**, 335-346.

Larin Z., Monaco A.P. and Lehrach H. (1991) Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci. USA,* **88**, 4123-4127.

Larsen F., Gunderson G., Lopez R. and Prydz H. (1992) CpG islands as gene markers in the human genome. *Genomics,* **13**, 1095-1107.

Leboyer M., Malafosse A., Boularand S., Campion D., Gheysen F., Samolyk D., Henriksson B., Denise E., des Lauriers A., Lepine J.P., Zarifan J-P., Clerget-Darpoux F. and Mallet J. (1990) Tyrosine hydroxylase polymorphisms associated with manic-depressive illness. *Lancet,* **335**, 1219.

Leff J. (1992) Over the edge: stress and schizophrenia. *New Scientist, 4th Jan,* 30-33.

Leonard S., Adler L.E., Bickford P.C., Hall M., Rollins Y., Breese C., Logel J., Drebing C., Adams C., Barnhart M., Byerley W., Coon H. and Freedman R. (1995) Association of neuronal nicotinic acetylcholine receptors with schizophrenia. *Schizophr. Res.* **15(1-2)**, 64

Levinson D.F. and Mowry B.J. (1991) Defining the schizophrenia spectrum:

issues for genetic linkage studies. *Schizophr. Bull.* **17(3)**, 491-514.

Levy D.L., Holzman P.S., Matthysse S. and Mendell N.R. (1994) Eye tracking and schizophrenia: a selective review. *Schizophr. Bull.* **20**, 47-62.

Levy-Lahad E., Wasco W., Poorkaj P., Romano D.M., Oshima J., Pettingell W.H., Yu C., Jondro P.D., Schmidt S.D., Wang K., Crowley A.C., Fu Y., Guenette S.Y., Galas D., Nemens E., Wijsman E.M., Bird T.D., Schellenburg G.D. and Tanzi R.E. (1995a) Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science,* **269**, 973-977.

Levy-Lahad E., Wijsman E.M., Nemens E., Anderson L., Goddard K.A.B., Weber J.L., Bird T.D. and Schellenburg G.D. (1995b) A familial Alzheimer's disease locus on chromosome 1. *Science,* **269**, 970-973.

Lim L.C., Gurling H., Curtis D., Brynjolfsson J., Petursson H. and Gill M. (1993) Linkage between tyrosine hydroxylase gene and affective disorder cannot be excluded in two of six pedigrees. *Am. J. Med. Genet.* **48**, 223-228.

Lim L.C., Nöthen M.M., Korner J., Rietschel M., Castle D., Hunt N., Propping P., Murray R. and Gill M. (1994) No evidence of association between dopamine D4 receptor variants and bipolar affective disorder. *Am. J. Med. Genet.* **54**, 259-263.

Lindor N.M., Sobell J.L., Heston L.L., Thibodeau S.N. and Sommer S.S. (1994) Screening the dystrophin gene suggests a high rate of polymorphism in general but no exonic deletions in schizophrenics. *Am. J. Med. Genet.* **54**, 1-4.

Lindsay S. and Bird A.P. (1987) Use of restriction enzymes to detect potential

gene sequences in mammalian DNA. *Nature,* **327**, 336-338.

Link A.J. and Olson M.V. (1991) Physical map of the *Saccharomyces cerevisiae* genome at 110kb resolution. *Genetics,* **127**, 681-698.

Litt M., Kramer P., Hauge X.Y., Weber J.L., Wang Z., Wilkie P.J., Holt M.S., Mishra S., Donis Keller H., Warnich L., Retief A.E., Jones C. and Weissenbach J. (1993) A microsatellite-based index map of human chromosome 11. *Hum. Mol. Genet.* **2**, 909-913.

Litt M. and Luty J.A. (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44**, 397-401.

Liu P., Legerski R. and Sicilliano M.J. (1989) Isolation of human transcribed sequences from human-rodent somatic cell hybrids. *Science,* **246**, 813-815.

Liu W.G. and Whittier R.F. (1995) Thermal Asymmetric Interlaced PCR: Automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics,* **25**, 674-681.

Livingston M.G. (1994) Risperidone. *Lancet,* **343**, 457-460.

Loh E.Y., Elliott J.F., Cwirla S., Lanier L.L. and Davis M.M. (1989) Polymerase chain reaction with single sided specificity: analysis of T cell receptor d chain. *Science,* **243**, 217-220.

Lohr J.B. and Bracha H.S. (1989) Can schizophrenia be related to prenatal exposure to alcohol? Some speculations. *Schizophr. Bull.* **15(4)**, 595-603.

Lohr J.B. and Bracha H.S. (1992) A monozygotic mirror-image twin pair with discordant psychiatric illnesses: a neuropsychiatric and neurodevelopmental evaluation. *Am. J. Psychiatry,* **149**, 1091-1095.

Lohr J.B. and Flynn K. (1992) Smoking and schizophrenia. *Schizophr. Res.* **8**, 93-102.

Lohr J.B. and Flynn K. (1993) Minor physical anomalies in schizophrenia and mood disorders. *Schizophr. Bull.* **19**, 551-556.

Lovett M. (1994) Fishing for complements: finding genes by direct selection. *Trends Genet.* **10**, 352-357.

Lovett M., Kere J. and Hinton L.M. (1991) Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. USA,* **88**, 9628-9632.

Macciardi F., Kennedy J.L., Ruocco L., Giuffra L., Carrera P., Marino C., Rinaldi V., Smeraldi E. and Ferrari M. (1992) A genetic linkage study of schizophrenia to chromosome 5 markers in a Northern Italian population. *Biol. Psychiatry,* **31**, 720-728.

Macciardi F., Verga M., Kennedy J.L., Petronis A., Bersani G., Pancheri P. and Smeraldi E. (1994) An association study between schizophrenia and the dopamine receptor genes DRD3 and DRD4 using haplotype relative risk. *Hum. Hered.* **44**, 328-336.

Mallet J., Meloni R. and Laurent C. (1994) Catecholamine metabolism and psychiatric or behavioral disorders. *Curr. Opin. Genet. Dev.* **4**, 419-426.

Mandelkow E. and Mandelkow E. (1995) Microtubules and microtubule-associated proteins. *Current Opinion in Cell Biology,* **7**, 72-81.

Mant R., Williams J., Asherson P., Parfitt E., McGuffin P. and Owen M.J. (1994) Relationship between homozygosity at the dopamine D3 receptor gene and schizophrenia. *Am. J. Med. Genet.* **54**, 21-26.

Marneros A., Deister A. and Rohde A. (1992) Comparison of long-term outcome of schizophrenic, affective and schizoaffective disorders. *Br. J. Psychiatry Suppl.* 44-51.

Matthysse S., Holzman P.S. and Lange K. (1986) The genetic transmission of schizophrenia: application of Mendelian latent strucutre analysis to eye tracking dysfunctions in schizophrenia and affective disorders. *J. Psychiatric Res.* **20**, 57-76.

Maule J.C. (1994) Electrophoretic karyotype analysis: pulsed field gel electrophoresis. In Gosden J.R. (ed.), *Chromosome Analysis Protocols.* Humana, Totowa, N.J. pp. 221-252.

McBride O.W. and Ozer H.L. (1973) Transfer of genetic information by purified metaphase chromosomes. *Proc. Natl. Acad. Sci. USA,* **70(4)**, 1258-1262.

McGrath J.J., Pemberton M.R., Welham J.L. and Murray R.M. (1994) Schizophrenia and the influenza epidemics of 1954, 1957 and 1959: a southern hemisphere study. *Schizophr. Res.* **14**, 1-8.

McGuffin P., Farmer A. and Gottesman I.I. (1987) Is there really a split in schizophrenia? The genetic evidence. *Br. J. Psychiatry,* **150**, 581-592.

McGuffin P., Reveley A. and Holland A. (1982) Identical triplets: non-identical psychosis? *Br. J. Psychiatry,* **140**, 1-6.

McGuffin P., Sargeant M., Hetti G., Tidmarsh S., Whatley S. and Marchbanks R.M. (1990) Exclusion of schizophrenia susceptibility gene from the chromosome 5q11-13 region: new data and a reanalysis of previous reports. *Am. J. Hum. Genet.* **47**, 524-535.

McGuffin P. and Stuart E. (1986) Genetic markers and schizophrenia. *Hum. Hered.* **36**, 65-88.

McNeil T.F., Cantor-Graae E. and Cardenal S. (1993) Prenatal cerebral development in individuals at genetic risk for psychosis: head size at birth in offspring of women with schizophrenia. *Schizophr. Res.* **10**, 1-5.

Mednick S.A., Machon R.A., Huttunen M.O. and Bonett D. (1988) Adult schizophrenia following prenatal exposure to an influenza epidemic. *Arch. Gen. Psychiatry,* **45**, 189-192.

Melmer G. and Buchwald M. (1992) Identification of genes using oligonucleotides corresponding to splice site consensus sequences. *Hum. Mol. Genet.* **1**, 433-438.

Mendlewicz J., Sevy S., Brocas H., Simon P., Charon F., Legros S. and Vassart G. (1987) Polymorphic DNA marker on X chromosome and manic depression. *Lancet,* **i**, 1230-1232.

Michels R. and Marzuk P.M. (1993) Progress in psychiatry (1). *N. Engl. J. Med.* **329**, 552-560.

Moon R.T. and McMahon A.P. (1990) Generation of diversity in nonerythroid spectrins. *J. Biol. Chem.* **265(8)**, 4427-4433.

Monaco A.P., Neve R.L., Colletti-Feener C., Bertelson C.J., Kurnit D.M. and Kunkel L.M. (1986) Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature,* **323**, 646-650.

Morgan J.G., Dolganov G.M., Robbins S.E., Hinton L.M. and Lovett M. (1992) The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. *Nucl. Acids Res.* **20**, 5173-5179.

Morris A.G., Gaitonde E., McKenna P.J., Mollon J.D. and Hunt D.M. (1995) *Hum. Mol. Genet.*, **4(10)**, 1957-1961.

Morton N.E. and McLean C.J. (1974) Analysis of family resemblance III. Complex segregation analysis of quantitative traits. *Am. J. Hum. Genet.* **26**, 489-503.

Mott F.W. (1911) Heredity and Insanity. *Lancet,* **3**, 1251-1259.

Mueller P.R. and Wold B. (1989) In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science,* **246**, 780-786.

Muir W.J., Gosden C.M., Brookes A.J., Fantes J., Evans K.L., Maguire S.M., Stevenson B., Boyle S., Blackwood D.H., St.Clair D.M., Porteous D.J. and Weith A. (1995) Direct microdissection and microcloning of a translocation breakpoint region, t(1;11)(q42.2;q21), associated with schizophrenia. *Cytogenet. Cell Genet.* **70**, 35-40.

Mullis K.B. and Faloona F.A. (1995) Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction. *Methods Enzymol.* **155**, 335-350.

Murray R.M., Lewis S. and Reveley A.M. (1985) Towards an etiological classification of schizophrenia. *Lancet,* **i**, 1023-1026.

Nanko S., Fukuda R., Hattori M., Sasaki T., Dai X.Y., Gill M., Kuwata S., Shibata Y. and Kazamatsuri H. (1994a) No evidence of linkage or allelic association of schizophrenia with DNA markers at pericentric region of chromosome 9. *Biol. Psychiatry,* **36**, 589-594.

Nanko S., Fukuda R., Hattori M., Sasaki T., Dai X.Y., Yamaguchi K. and Kazamatsuri H. (1994b) Further evidence of no linkage between schizophrenia and the dopamine D3 receptor gene locus. *Am. J. Med. Genet.* **54**, 264-267.

Nanko S., Gill S., Owen M., Takazawa N., Moridaiwa J. and Kazamatsuri H. (1992) Linkage study of schizophrenia with markers on chromosome 11 in two Japanese pedigrees. *Jap. J. Psychiatr. Neurol.* **46(1)**, 155-159.

Nanko S., Hattori M., Dai X.Y., Fukuda R. and Kazamatsuri H. (1994c) DRD2 Ser311/Cys311 polymorphism in schizophrenia. *Lancet,* **343**, 1044

Nanko S., Kunugi H., Sasaki T., Fukuda R., Kawate T. and Kazamatsuri H. (1993a) Pericentric region of chromosome 9 is a possible candidate region for linkage study of schizophrenia. *Biol. Psychiatry,* **33**, 655-658.

Nanko S., Sasaki T., Fukuda R., Hattori M., Dai X.Y., Kazamatsuri H., Kuwata S., Juji T. and Gill M. (1993b) A study of the association between schizophrenia and the dopamine D3 receptor gene. *Hum. Genet.* **92**, 336-338.

Nelson D.L., Ledbetter S.A., Corbo L., Vicotia M.F., Ramirez-Solis R., Webster T.D., Ledbetter D.H. and Caskey C.T. (1989) Alu polymerase chain reaction: a method for rapid amplification of human specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. USA,* **86**, 1-5.

Nelson W.J. and Veshnock P.J. (1987) Modulation of fodrin (membrane skeleton) stability by cell-cell contact in Madin-Darby canine kidney epithelial cells. *J.Cell. Biol.*, **10**4, 1527-1537.

Neylan T.C. and van Kammen D.P. (1990) Biological mechanisms of schizophrenia: an update. *Psychiatr. Med.* **8**, 41-52.

NIH/CEPH Collaborative Mapping Group (1992) A comprehensive linkage map of the human genome. *Science,* **258**, 67-87.

Nimgaonkar V.L., Scott J.A., Brar J.S., Ganguli R. and Chakravarti A. (1993a) Co-occurrence of schizophrenia and Treacher Collins syndrome. *Am. J. Med. Genet.* **48**, 156-158.

Nimgaonkar V.L., Zhang X.R., Caldwell J.G., Ganguli R. and Chakravarti A. (1993b) Association study of schizophrenia with dopamine D3 receptor gene polymorphisms: probable effects of family history of schizophrenia? *Am. J. Med. Genet.* **48**, 214-217.

Ninomiya Y., Gordon M., van der Rest M., Schmid T., Linsenmayer T. and Olsen B.R. (1995) The developmentally regulated type X collagen gene contains a long open reading frame without introns. *J. Biol. Chem.* **261**, 5041-5050.

Nöthen M.M., Cichon S., Hemmer S., Hebebrand J., Remschmidt H., Lehmkuhl G., Poustka F., Schmidt M., Catalano M., Fimmers R., Körner J., Rietschel M. and Propping P. (1994) Human dopamine D4 receptor gene: frequent occurrence of a null allele and observation of homozygosity. *Hum. Mol. Genet.* **3**, 2207-2212.

Nöthen M.M., Cichon S., Propping P., Fimmers R., Schwab S.G. and Wildenauer D.B. (1993) Excess of homozygosity at the dopamine D3 receptor gene in schizophrenia not confirmed. *J. Med. Genet.* **30**, 708-709.

O'Callaghan E., Sham P., Takei N., Glover G. and Murray R.M. (1991) Schizophrenia after prenatal exposure to 1957 A2 influenza epidemic. *Lancet,* **337**, 1248-1249.

O'Callaghan E., Sham P.C., Takei N., Murray G., Glover G., Hare E.H. and Murray R.M. (1994) The relationship of schizophrenic births to 16 infectious diseases. *Br. J. Psychiatry,* **165**, 353-356.

O'Donovan M.C., Guy C., Craddock N., Murphy K.C., Cardno A.G., Jones L.A., Owen M.J. and McGuffin P. (1995) Expanded CAG repeats in schizophrenia and bipolar disorder. *Nat. Genet.,* **10**, 380-381.

O'Neil J., Brynjolfsson J., Curtis D., Rifkin L., Maloney E., Murphy P., Petursson H. and Gurling H. (1993) Testing the hypothesis for genomic imprinting and anticipation in bipolar and related unipolar affective disorder. *Psych. Genet.* **3**, 151

O'Reilly R.L. and Davis B.A. (1994) Phenylethylamine and schizophrenia. *Prog. Neuropsychopharmacol. & Biol. Psychiat.* **18**, 63-75.

Ochman H., Gerber A.S. and Hartl D.L. (1988) Genetic applications for an inverse polymerase chain reaction. *Genetics,* **120**, 621-623.

Olson M., Hood L., Cantor C. and Botstein D. (1989) A common language for the physical mapping of the human genome. *Science,* **245**, 1434-1435.

Orbach M.J., Vollrath D., Davis R.W. and Yanofsky C. (1988) An electrophoretic karyoptype of *Neurospora crassa*. *Mol. Cell. Biol.* **8(4)**, 1469-1473.

Owen F. and Simpson M. (1994) The neurochemistry of schizophrenia. In Owen F. and Itzhaki R. (eds.), *Molecular and Cell Biology of Neuropsychiatric Diseases.* Chapman & Hall, London, pp. 133-159.

Owen M.J., Mant R., Parfitt E., Williams J., Asherson P., O'Mahoney G., Van Os J., Llewellyn D., Collier D., Gill M. and McGuffin P. (1992) No association between RFLPs at the porphobilinogen deaminase gene and schizophrenia. *Hum. Genet.* **90**, 131-132.

Parfitt E., Asherson P., Sargeant M., Whatley S., McGuffin P. and Owen M. (1991) A linkage study of the pseudoautosomal region in schizophrenia. *Psychiatr. Genet.* **2**, 92-93.

Parimoo S., Patanjali S.R., Shukla H., Chaplin D.D. and Weissman S.M. (1991) cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci. USA,* **88**, 9623-9627.

Parker J.D., Rabinovitch P.S. and Burmer G.C. (1991) Targeted gene walking polymerase chain reaction. *Nucl. Acids Res.* **19**, 3055-3060.

Parks C.L., Chang L. and Shenk T. (1991) Polymerase chain reaction mediated by a single primer: cloning of genomic sequences adjacent to a serotonin receptor protein coding region. *Nucl. Acids Res.* **19**, 7155-7160.

Paschal B.M., Obar R.A. and Vallee R.B. (1989) Interaction of brain cytoplasmic dynein and MAP2 with a common sequence at the C terminus of tubulin. *Nature*, **342**, 569-572.

Patel K., Sheer D. and Hampton G.M. (1993) "Junction Trapping". A simple PCR-based method for the isolation of YAC-insert termini. *GATA,* **10**, 42-48.

Pearson W.R. and Lipman D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA,* **85**, 2444-2448.

Persico A.M., Wang Z.W., Black D.W., Andreasen N.C., Uhl G.R. and Crowe R.R. (1995) Exclusion of close linkage of the dopamine transporter gene with schizophrenia spectrum disorders. *Am. J. Psychiatry,* **152**, 134-136.

Petronis A. and Kennedy J.L. (1995) Unstable genes - unstable mind? *Am. J. Psychiatry,* 164-172.

Pickar D., Owen R.R., Jr., Litman R.E., Hsiao J.K. and Su T.P. (1994) Predictors of clozapine response in schizophrenia. *J. Clin. Psychiatry,* **55 Suppl B**, 129-132.

Polymeropoulos M.H., Coon H., Byerley W., Gershon E.S., Goldin L., Crow T.J., Rubenstein J., Hoff M., Holik J., Smith A.M., Shields G., Bass N.J., Poulter M., Lofthouse R., Vita A., Morganti C., Merril C.R. and DeLisi L.E. (1994) Search for

a schizophrenia susceptibility locus on human chromosome 22. *Am. J. Med. Genet.* **54**, 93-99.

Polymeropoulos M.H., Xiao H., Glodek A., Gorski M., Adams M.D., Moreno R.F., Fitzgerald M.G., Venter J.C. and Merril C.R. (1992) Chromosomal assignment of 46 brain cDNAs. *Genomics,* **12**, 492-496.

Polymeropoulos M.H., Xiao H., Sikela J.M., Adams M., Venter J.C. and Merrill C.R. (1993) Chromosomal distribution of 320 genes from a brain cDNA library. *Nat. Genet.* **4**, 381-386.

Pope H.G., Jonas J.M., Cohen B.M. and Lipinski J.F. (1982) Failure to find evidence of schizophrenia in first degreee relatives of schizophrenic probands. *Am. J. Psychiatry,* **139**, 826-827.

Potts N.L.S., Davidson J.R.T. and Krishnan K.R.R. (1993) The role of nuclear magnetic resonance imaging in psychiatric research. *J. Clin. Psychiatry,* **54:12** (suppl), 13-18.

Prescott C.A. and Gottesman I.I. (1993) Genetically mediated vulnerability to schizophrenia. *Psychiatr. Clin. North Am.* **16**, 245-267.

Pulver A.E., Karayiorgou M., Lasseter V.K., Wolyniec P., Kasch L., Antonarakis S., Housman D., Kazazian H.H., Meyers D., Nestadt G., Ott J., Liang K-Y., Lamacz M., Thomas M. and Childs B. (1994b) Follow-up of a report of a potential linkage for schizophrenia on chromosome 22q12-q13.1: Part 2. *Am. J. Med. Genet.* **54**, 44-50.

Pulver A.E., Karayiorgou M., Wolyniec P.S., Lasseter V.K., Kasch L., Nestadt

G., Antonarakis S., Housman D., Kazazian H.H., Meyers D., Ott J., Lamacz M., Liang K-Y., Hanfelt J., Ullrich G., Dimarchi N., Ramu E., McHugh P.M., Adler L., Thomas M., Carpenter W.T., Mansschreck T., Gordon C.T., Kimberland M., Babb R., Puck J. and Childs B. (1994a) Sequential strategy to identify a susceptibility gene for schizophrenia: report of potential linkage on chromosome 22q12-q13.1: Part 1. *Am. J. Med. Genet.* **54**, 36-43.

Pulver A.E., Liang K.Y., Brown C.H., Wolyniec P., McGrath J., Adler L., Tam D., Carpenter W.T. and Childs B. (1992) Risk factors in schizophrenia. Season of birth, gender, and familial risk. *Br. J. Psychiatry,* **160**, 65-71.

Rao M.L. and Moller H.J. (1994) Biochemical findings of negative symptoms in schizophrenia and their putative relevance to pharmacologic treatment. A review. *Neuropsychobiol.,* **30**, 160-172.

Rasenick M.M., Wang N. and Yan K. (1990) Specific association between tubulin and G proteins: participation of cytoskeletal elements in cellular signal transduction. *Adv. Second Mess. & Phos. Res.*, **24**, 381-386.

Reveley A.M., Reveley M.A., Clifford C.A. and Murray R.M. (1982) Cerebral ventricular size in twins disocrdant for schizophrenia. *Lancet,* **i**, 540-541.

Ridley R.M. and Baker H.F. (1990) Implications of age of onset for the genetics of schizophrenia. *Biol. Psychiatry,* **28**, 455-458.

Riecher Rössler A., Hafner H., Dutsch Strobel A., Oster M., Stumbaum M., van Gulick Bailer M. and Loffler W. (1994) Further evidence for a specific role of estradiol in schizophrenia? *Biol. Psychiatry,* **36**, 492-494.

Rieder R.O. and Gershon E.S. (1978) Genetic strategies in biological psychiatry. *Arch. Gen. Psychiatry,* **35**, 866-873.

Rifkin L., Lewis S., Jones P., Toone B. and Murray R. (1994) Low birth weight and schizophrenia. *Br. J. Psychiatry,* **165**, 357-362.

Riley J., Butler R., Ogilvie D., Finniear R., Jenner D., Powell S., Anand R., Smith J.C. and Markham A.F. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucl. Acids Res.* **18**, 2887-2890.

Risch N. (1992) Genetic linkage: interpreting lod scores. *Science,* **255**, 803-804.

Roberts G.W. and Bruton C.J. (1990) Notes from the graveyard: neuropathology and schizophrenia. *Neuropathol. Appl. Neurobiol.* **16**, 3-16.

Rodionov V.I., Gyoeva F.K., Kashina A.S., Kuznetsov S.A. and Gelfand V.I. (1990) Microtubule-associated proteins and microtubule-based translocators have different binding sites on tubulin molecule. *J. Biol. Chem.* **265(10)**, 5702-5707.

Rogaev E.I., Sherrington R., Rogaeva E.A., Levesque G., Ikeda M., Llang Y., Chi H., Lin C., Holman K., Tsuda T., Mar L., Sorbi S., Nacmias B., Piacentini S., Amaducci L., Chumakov I., Cohen D., Lannfelt L., Fraser P.E., Rommens J.M. and St.George-Hyslop P.H. (1995) Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature,* **376**, 775-778.

Rosenthal A. and Jones D.S.C. (1990) Genomic walking and sequencing by

oligo-cassette mediated polymerase chain reaction. *Nucl. Acids Res.* **18**, 3095-3096.

Rosenthal D., Wender P.H., Kety S.S., Welner J. and Schulsinger F. (1971) The adopted-away offspring of schizophrenics. *Am. J. Psychiatry,* **128**, 307-311.

Roth B.L. (1994) Multiple serotonin receptors: clinical and experimental aspects. *Ann. Clin. Psychiatry,* **6**, 67-78.

Roux K.H. and Dhanarajan P. (1990) A strategy for single site PCR amplification of dsDNA: priming digested cloned or genomic DNA from an anchor-modified restriction site and a short internal sequence. *Biotechniques,* **8**, 48-57.

Roychowdhury S. and Rasenick M.M. (1994) Tubulin-G protein association stabilises GTP-binding and activates GTPase: cytoskeletal participation in neuronal signal transduction. *Biochemistry,* **33**, 9800-9805.

Rubinsztein D.C., Leggo J., Goodburn S., Crow T.J., Lofthouse R., DeLisi L.E., Barton D.E. and Ferguson Smith M.A. (1994) Study of the Huntington's disease (HD) gene CAG repeats in schizophrenic patients shows overlap of the normal and HD affected ranges but absence of correlation with schizophrenia. *J. Med. Genet.* **31**, 690-693.

Rybakowski J.K. and Lehmann W. (1994) Decreased activity of erythrocyte membrane ATPases in depression and schizophrenia. *Neuropsychobiology,* **30**, 11-14.

Saiki R.K., Gelfand D.H., Stoffel S., Schart S.J., Higuchi R., Horn G.T., Mullis

K.B. and Erlich H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science,* **239**, 487-491.

Sambrook J., Fritsch E.F. and Maniatis T. (1989) *Molecular Cloning. A laboratory manual.* CSH Laboratory Press, Cold Spring Harbor.

Sanders A.R., Hamilton J.D., Fann W.E. and Patel P.I. (1991) Association of genetic variation in the porphobilinogen deaminase gene with schizophrenia. *Am. J. Hum. Genet.* **49**, A2011

Sarkar G., Kapelner S., Grandy D.K., Marchionni M., Civelli O., Sobell J., Heston L. and Sommer S.S. (1991) Direct sequencing of the dopamine D2 receptor (DRD2) in schizophrenics reveals three polymorphisms but no structural change in the receptor. *Genomics,* **11**, 8-14.

Sasaoka T. and Tonegawa S. (1995) A molecular genetic approach to the pathogenesis of schizophrenia using genetically engineered mice. *Schizophr. Res.* **15(1-2)**, 70.

Schwartz D.C. and Cantor C.R. (1984) Separation of yeast chromosome-sized DNAs by pulsed field gel electrophoresis. *Cell,* **37**, 67-75.

Seeman P., Guan H.C. and Van Tol H.H.M. (1994a) Dopamine D4 receptors elevated in schizophrenia. *Nature,* **365**, 441-445.

Seeman P., Ulpian C., Chouinard G., Van Tol H.H., Dwosh H., Lieberman J.A., Siminovitch K., Liu I.S., Waye J., Voruganti P., Hudson C., Serjeant G.R., Masibay A.S. and Seeman M.V. (1994b) Dopamine D4 receptor variant,

D4GLYCINE194, in Africans, but not in Caucasians: no association with schizophrenia. *Am. J. Med. Genet.* **54**, 384-390.

Selkoe, D.J. (1995) Missense on the membrane. *Nature*, **375**, 734-735.

Selten J.P. and Slaets J.P. (1994) Evidence against maternal influenza as a risk factor for schizophrenia. *Br. J. Psychiatry,* **164**, 674-676.

Senger G., Ludecke H.J., Horsthemke B. and Claussen U. (1990) Microdissection of banded human chromosomes. *Hum. Genet.* **84**, 507-511.

Serra A., Brahe C., Millington Ward A., Neri G., Tedeschi B., Tassone F. and Bova R. (1990) Pericentric inversion of chromosome 9: prevalence in 300 Down syndrome families and molecular studies of nondisjunction. *Am. J. Med. Genet. Suppl.* **7**, 162-168.

Shaikh S., Ball D., Craddock N., Castle D., Hunt N., Mant R., Owen M., Collier D. and Gill M. (1993) The dopamine D3 receptor gene: no association with bipolar affective disorder. *J. Med. Genet.* **30**, 308-309.

Shaikh S., Gill M., Owen M., Asherson P., McGuffin P., Nanko S., Murray R.M. and Collier D.A. (1994) Failure to find linkage between a functional polymorphism in the dopamine D4 receptor gene and schizophrenia. *Am. J. Med. Genet.* **54**, 8-11.

Shapiro M.B. and Senapathy P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* **15(17)**, 7155-7174.

Sharp C.W., Muir W.J., Blackwood D.H.R., Gosden C. and St.Clair D.M. (1994) Schizophrenia and mental retardation associated in a pedigree with retinitis pigmentosa and sensorineural deafness. *Am. J. Med. Gen.* **54**, 354-360.

Shen M.R., Batzer M.A. and Deininger P.L. (1991) Evolution of the master Alu gene(s). *J. Mol. Evol.* **33**, 311-320.

Shen W.W. (1994) Pharmacotherapy of schizophrenia: the American current status. *Keio. J. Med.* **43**, 192-200.

Sherrington R., Brynjolfsson J., Petursson H., Potter M., Dudleston K., Barraclough B., Wasmuth J., Dobbs M. and Gurling H. (1988) Localization of a susceptibility locus for schizophrenia on chromosome 5. *Nature,* **336**, 164-167.

Shibasaki Y., Maule J.C., Devon R.S., Slorach E.M., Gosden J.R., Porteous D.J. and Brookes A.J. (1995) Catch-linker + PCR labeling: a simple method to generate fluorescence in situ hybridisation probes from yeast artificial chromosomes. *PCR Methods and Applications,* **4**, 209-211.

Shields J. and Slater E. (1967) Genetic aspects of schizophrenia. *Hospital Medicine,* 579-584.

Shizuya H., Birren B., Kim U., Mancino V., Slepak T., Tachiri Y. and Simon M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia Coli using an F-factor based vector. *Proc. Natl. Acad. Sci. USA,* **89**, 8794-8797.

Siebert P.D., Chenchik A., Kellogg D.E., Lukyanov K.A. and Lukyanov S.A.

(1995) An improved PCR method for walking in uncloned genomic DNA. *Nucl. Acids Res.* **23**, 1087-1088.

Smith M., Wasmuth W., McPherson J.D., Wagner C., Grandy D., Civelli O., Potkin S. and Litt M. (1989) Cosegregation of an 1q22-9p22 translocation with affective disorder: proximity of the dopamine D2 receptor gene relative to the translocation breakpoint. *Am. J. Hum. Genet.* **45**, A220

Smith M.W., Holmsen L.N., Wei Y.H., Peterson M. and Evans G.A. (1994) Genomic sequence sampling: a strategy for high resolution sequence-based physical mapping of complex genomes. *Nat. Genet.* **7**, 40-47.

Sobell J.L., Heston L.L. and Sommer S.S. (1993) Novel association approach for determining the genetic predisposition to schizophrenia: case-control resource and testing of a candidate gene. *Am. J. Med. Genet.* **48**, 28-35.

Sobue K. (1993) Actin-based cytoskeleton in growth cone activity. *Neurosci. Res.* **18**, 91-102.

Sokoloff P., Giros B., Martres M.P., Bouthenet M.L. and Schwartz J.C. (1990) Molecular cloning and characterization of a novel dopamine receptor (D3) as a target for neuroleptics. *Nature,* **347**, 146-151.

Sommer R. and Tautz D. (1989) Minimal homology requirements for PCR primers. *Nucl. Acids Res.* **17**, 6749.

Sommer S.S., Lind T.J., Heston L.L. and Sobell J.L. (1993) Dopamine D4 receptor variants in unrelated schizophrenic cases and controls. *Am. J. Med. Genet.* **48**, 90-93.

Southern E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.

Squires R.F. and Saederup E. (1991) A review of evidence for GABergic predominance/glutamatergic deficit as a common etiological factor in both schizophrenia and affective psychoses: more support for a continuum hypothesis of "functional" psychosis. *Neurochem. Res.* **16**, 1099-1111.

St.Clair D., Blackwood D., Muir W., Carothers A., Walker M., Spowart G., Gosden C. and Evans H.J. (1990) Association within a family of a balanced autosomal translocation with major mental illness. *Lancet,* **336**, 13-16.

St.Clair D. (1994) Expanded CAG trinucleotide repeat of Huntington's disease gene in a patient with schizophrenia and normal striatal histology. *J. Med. Genet.* **31**, 658-659.

St.Clair D., Blackwood D., Muir W., Baillie D., Hubbard A., Wright A. and Evans H.J. (1989) No linkage of chromosome 5q11-q13 markers ot schizophrenia in Scottish families. *Nature,* **339**, 305-309.

St.Clair D., Muir W. and Blackwood D. (1994) Molecular Biology of Schizophrenia. In Owen F. and Itzhaki R. (eds.), *Molecular and Cell Biology of Neuropsychiatric Diseases.* Chapman and Hall, London, pp. 160-172.

Stein G.S., Stein J.L. and Marzluff W.F. (1995) *Histone genes: structure, organization and regulation.* John Wiley and Sons, Inc. New York.

Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R., Halloran N., Green P., Thierry Mieg J., Qiu L., Dear S., Coulson A., Craxton M., Durbin R.,

Berks M., Metzstein M., Hawkins T., Ainscough R. and Waterston R. (1992) The C. elegans genome sequencing project: a beginning. *Nature,* **356**, 37-41.

Sunahara R.K., Niznik H.B., Weiner D.M., Stormann T.M., Brann M.R., Kennedy J.L., Gelernter J.E., Rozmahel R., Yang Y.L., Israel Y., Seeman P. and O'Dowd B.F. (1990) Human dopamine D1 receptor encoded by an intronless gene on chromosome 5. *Nature,* **347**, 80-83.

Susser E., Neugebauer R., Hoek H., Brown A., Lin S., Labowitz D. and Gorman J. (1995) An epidemiologic study of schizophrenia after prenatal exposure to famine. *Schizophr. Res.* **15(1-2)**, 197.

Sutcliffe J.G. (1988) mRNA in the mammalian central nervous system. *Ann. Rev. Neurosci.* **11**, 157.

Szpirer C., Molne M., Antonacci R., Jenkins N.A., Finelli P., Szpirer J., Riviere M., Rocchi M., Gilbert D.J., Copeland N.G. and Gallo V. (1994) The genes encoding the glutamate receptor subunits KA1 and KA2 (GRIK4 and GRIK5) are located on separate chromosomes in human, mouse, and rat. *Proc. Natl. Acad. Sci. USA,* **91**, 11849-11853.

Tadokoro K., Oki N., Fujii H., Oshima A., Inoue T. and Yamada M. (1995) Genomic organisation of the human WT1 gene. *Jpn. J. Cancer Res.* **83**, 1198-1203.

Tassone F., Xu H., Burkin H., Weissman S. and Gardiner K. (1995) cDNA selection from 10Mb of chromosome 21 DNA: efficiency in transcriptional mapping and reflections of genome organisation. *Hum. Mol. Gen.* **4(9)**, 1509-1518.

Taylor G.R. (1991) Polymerase chain reaction: basic principles and automation. In McPherson M.J., Quirke P. and Taylor G.R. (eds.), *PCR. A Practical Approach.* Oxford University Press, Oxford, Vol.1. pp. 1-14.

Terwilliger J.D. and Ott J. (1994) Handbook of Human Genetic Linkage. The Johns Hopkins University Press, Baltimore and London.

Tienari P., Wynne L.C., Moring J., Lahti I., Naarala M., Sorri A., Wahlberg K.E., Saarento O., Seitamaa M., Kaleva M. and Läksy K. (1994) The Finnish adoptive family study of schizophrenia. Implications for family research. *Br. J. Psychiatry Suppl.* 20-26.

Todd S. and Naylor S.L. (1991) Dinucleotide repeat polymorphism in the human alpha tubulin 1 (testis specific) gene (TUBA1). *Nucl. Acids Res.* **19**, 3755.

Torrey E.F., Bowler A.E. and Rawlings R. (1991) An influenza epidemic and the seasonality of schizophrenic births. In Kurstak E. (ed.), *Second World Congress on Viruses and Mental Health.* Plenum Press, New York, pp. 109-116.

Torrey W.C. and Drake R.E. (1994) Current concepts in the treatment of schizophrenia. *Psychiatry,* **57**, 278-285.

Trofatter J.A., MacCollin M.M., Rutter J.L., Murrell J.R., Duyao M.P., Parry D.M., Eldridge R., Kley N., Menon A.G., Pulaski K., Haase V.H., Ambrose C.M., Munroe D., Bove C., Haines J.L., Martuza R.L., MacDonald M.E., Seizinger B.R., Short M.P., Buckler A.J. and Gusella J.F. (1993) A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. *Cell,* **75**, 826.

Upcroft P. and Healey A. (1991) PCR priming from the restriction endonuclease site 3' extension. *Nucl. Acids Res.* **21**, 4854.

Upender M., Gallagher P.G., Moon R.T. and Forget B.T. (1994) Localisation of the human alpha-fodrin gene (SPTAN1) to 9q33-q34 by fluorescence in situ hybridisation. *Cytogenet. Cell. Genet.* **66**, 39-41.

Valdes J.M., Tagle D.A. and Collins F.S. (1994) Island rescue PCR: a rapid and efficient method for isolating transcribed sequences from yeast artificial chromosomes and cosmids. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5377-5381.

Van Tol H.H.M., Bunzow J.R., Guna H.C., Sunahara R.K., Seeman P., Niznik H.B. and Jovanovic V. (1991) Cloning the gene for a human dopamine D4 receptor with high affinity for the antipsychotic clozapine. *Nature,* **350**, 610-614.

Van Tol H.H.M., Wu C.M., Guan H.C., Ohara K., Bunzow J.R., Civelli O., Kennedy J., Seeman P., Niznik H.B. and Jovanovic V. (1992) Multiple dopamine D4 receptor variants in the human population. *Nature,* **358**, 149-152.

Waddington J.L. (1993) Schizophrenia: developmental neuroscience and pathobiology. *Lancet,* **341**, 531-536.

Waddington J.L., O'Callaghan E., Larkin C. and Kinsella A. (1993) Cognitive dysfunction in schizophrenia: organic vulnerability factor or state marker for tardive dyskinesia? *Brain Cogn.* **23**, 56-70.

Wallace M. (1987) The forgotten illness, '*The Times' newspaper*, May.

Wang S., Sun C., Walczak C.A., Ziegle J.S., Kipps B.R., Goldin L.R. and Diehl S.R. (1995) Evidence for a susceptibility locus for schizophrenia on chromosome 6pter-p22. *Nat. Genet.* **10**, 41-46.

Weber J.L. and May P.E. (1989) Abundant class of human DNA polmorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388-396.

Weil D., Blanchard S., Kaplan J., Guilford P., Gibson F., Walsh J., Mburu P., Varela A., Levilliers J., Weston M.D., Kelley P.M., Kimberling W.J., Wagenaar M., Levi-Acobas F., Larget-Piet D., Munnich A., Steel K.P., Brown S.D.M., and Petit C. (1995) Defective myosin VIIA gene responsible for Usher syndrome type 1B. *Nature,* **374**, 60-61.

Weiner H. (1995) Schizophrenia: etiology. In Kaplan H. and Sadock I. (eds.), *Comprehensive Textbook of Psychiatry.* Balti, Williams and Wilkins, Vol.4th. pp. 650-679.

Weiss M. and Green H. (1967) Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. *Proc. Natl. Acad. Sci. USA,* **58**, 1104-1111.

Weissenbach J., Gyapay G., Dib C., Vignal A., Morissette J., Millasseau P., Vaysseix G. and Lathrop M. (1992) A second-generation linkage map of the human genome. *Nature,* **359**, 794-801.

Weith A. (1994) The analysis of disease-specific chromosomal rearrangements with microcloned DNA markers. In Adolph K. (ed.), *Methods in Molecular Genetics, Vol 5, Part C.* Academic Press, London, pp169-194.

Welham J.L., McLachlan G.J. and McGrath J.J. (1995a) Time-series as an alternative analysis of seasonality in schizophrenia birth-rates. *Schizophr. Res.* **15(1-2)**, 202.

Welham J.L., Vacca J. and McGrath J.J. (1995b) Seasonality in schizophrenia births: a review of Southern hemisphere studies. *Schizophr. Res.* **15(1-2)**, 202.

Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N., Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R., Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J. and Wohldman P. (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans. *Nature,* **368**, 32-38.

Winship P.R. (1989) An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucl. Acids Res.* **19**, 1266-1226.

Wright P., Gill M. and Murray R.M. (1993) Schizophrenia: genetics and the maternal immune response to viral infection. *Am. J. Med. Genet.* **48**, 40-46.

Wyatt R.J., Alexander R.C., Egan M.F. and Kirch D.G. (1988) Schizophrenia, just the facts. What do we know, how well do we know it? *Schizophr. Res.* **1**, 3-18.

Yang L., Li T., Wiese C., Lannfelt L., Sokoloff P., Xu C.T., Zeng Z., Schwartz

J.C., Liu X. and Moises H.W. (1993) No association between schizophrenia and homozygosity at the D3 dopamine receptor gene. *Am. J. Med. Genet.* **48**, 83-86.

Yanisch-Perron C., Vieira J. and Messing J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. *Gene,* **33**, 103-119.

Yon S., Palmer R.W., Sheer D. and Fried M. (1989) Localisation of the surfeit gene cluster containing the ribosomal protein L7a to chromosome bands 9q33-34. *Ann. Hum. Genet.* **53**, 149-155.

Zamani M.G., De Hert M., Spaepen M., Hermans M., Marynen P., Cassiman J.J. and Peuskens J. (1994) Study of the possible association of HLA class II, CD4, and CD3 polymorphisms with schizophrenia. *Am. J. Med. Genet.* **54**, 372-377.

Zatz M., Vallada H., Melo M.S., Passos Bueno M.R., Vieira A.H., Vainzof M., Gill M. and Gentil V. (1993) Cosegregation of schizophrenia with Becker muscular dystrophy: susceptibility locus for schizophrenia at Xp21 or an effect of the dystrophin gene in the brain? *J. Med. Genet.* **30**, 131-134.

Zheng Y., Jung M.K. and Oakley B.R. (1991) Gamma-tubulin is present in Drosophila melanogaster and Homo sapiens and is associated with the centrosome. *Cell,* **65**, 817-823.

PAPERS PRESENTED DURING THE COURSE OF THIS THESIS

1644–1645 Nucleic Acids Research, 1995, Vol. 23, No. 9

Splinkerettes—improved vectorettes for greater efficiency in PCR walking

Rebecca S. Devon*, David J. Porteous and Anthony J. Brookes

MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

Received October 3, 1994; Revised and Accepted November 25, 1994

Vectorettes enable PCR amplification of DNA sequences which lie between a single known primer and a nearby restriction site. They have been applied in the isolation of end fragments from YAC recombinants (1) and also in PCR walking (2). Vectorettes consist of a double-stranded linker sequence with a central region of mismatch and a cohesive end suitable for ligation to restriction enzyme-digested DNA (Fig. 1). The vectorette primer employed in the PCR is exactly the same sequence as the mismatched portion of the upper strand, and therefore cannot anneal to and initiate priming from the vectorette until its complementary sequence has been synthesised by polymerase extension from the specific target DNA primer.

The specificity of the vectorette PCR reaction is, however, not absolute. Illegitimate products result from non-specific annealing of either primer and from a reaction we term 'end-repair' priming. End-repair priming involves the free cohesive ends of unligated vectorettes and inserts which are based on restriction sites that produce 5′ overhangs. These ends are filled in during the first cycle of the PCR reaction. After the subsequent denaturing step, these ends are able to anneal to each other with sufficient stability to initiate priming (3). Extension across a vectorette sequence at either end of the insert molecule will result in the production of a sequence complementary to the vectorette primer, allowing exponential PCR amplification without involvement of the specific target DNA primer. Vectorette dimers will also be produced by end-repair priming, but will not be amplified in the PCR because the two halves will form a stable hairpin structure which will be refractory to amplification.

The splinkerette we describe here is designed to decrease end-repair priming. Rather than a central DNA mismatch, the splinkerette incorporates a hairpin structure on the bottom strand (Fig. 1). The primer is of the same sequence as the top strand and therefore, as with the vectorette, is unable to act as a primer until the complement of this strand has been synthesised. In a PCR reaction the free 3′ end of the bottom strand will flip back on itself to form a hairpin and begin elongation further along the bottom strand. The resulting double-stranded structure is stable and is functionally removed from further reaction. It is therefore not able to cause end-repair priming. Furthermore, in the splinkerette system, only the top strand is available to act as a non-specific primer whereas in the vectorette system both the top strand (which after end-repair is four bases longer than that of the splinkerette) and the bottom strand could cause mispriming in this way. Splinkerettes are not kinased, and as such there is no covalent bond between the splinkerette bottom strand and the DNA to be amplified. This avoids elongation from the hairpin along the whole length of the insert molecule.

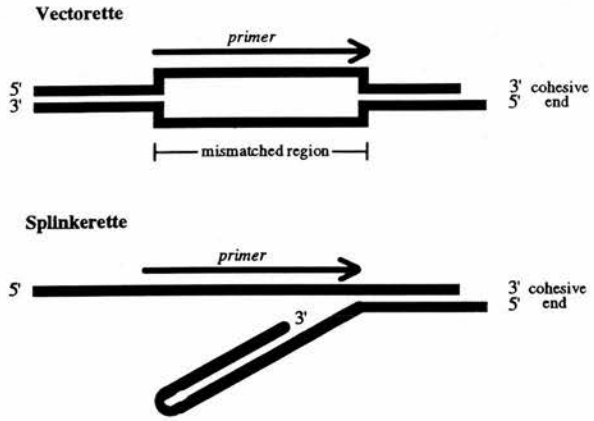


Figure 1. Schematic diagram of the vectorette and splinkerette.

Here we demonstrate the practical advantages of splinkerettes over vectorettes by comparing the efficiency of both in producing unique products from the complex resource of total human genomic DNA. To enable direct comparison, splinkerettes and vectorettes were designed to be compatible with a single primer.

For small target PCR products ($\leq$500 bp) it was found that the efficiency of both the splinkerette and vectorette was adequate to produce the expected products after one or two rounds of PCR (data not shown). However, Figure 2 shows the advantages of splinkerettes over vectorettes in the amplification of larger fragments where formation of the target product may be hindered by increased competition from end-repair primed artefacts. A 2.3 kb fragment from the WT1 gene (4) was successfully amplified only with the splinkerette. We ascribe the reproducible failure of vectorette PCR to amplify this fragment, and the smear of non-specific products instead resulting, to end-repair priming. The usefulness of splinkerette PCR has been further demonstrated by the routine isolation of long end fragments (up to 3.6 kb) from YAC recombinants (5).

* To whom correspondence should be addressed

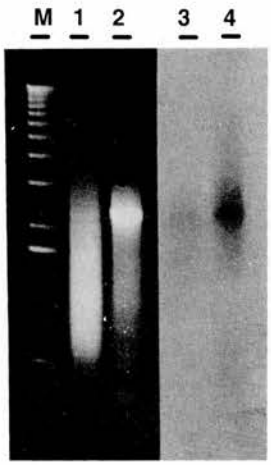


Figure 2. Comparison of splinkerettes and vectorettes in PCR walking. The target PCR product was a 2.3 kb fragment derived from the WTI gene (3). Vectorettes were obtained from ICI. Splinkerettes were made by duplexing the top strand (5′-CGA ATC GTA ACC GTT CGT ACG AGA ATT CGT ACG AGA ATC GCT GTC CTC TCC AAC GAG CCA AGG-3′) and the bottom strand (5′-GAT CCC TTG GCT CGT TTT TTT TTG CAA AAA-3′) by mixing the oligonucleotides (150 μg/ml each) at 90°C in 10 mM Tris–HCl (pH 7.4), 5 mM $MgCl_2$ and leaving to cool to room temperature. Total human genomic DNA was digested with *Bam*HI and ligated with a 15 × molar excess of splinkerettes or vectorettes in 20 μl for 4½ h at room temperature. PCR combined Hot Start (6) and Touchdown (7) protocols and conditions were as follows: denaturation, 95°C for 30 s in the first cycle and 15 s thereafter; annealing, for 1 min at 71°C initially, decreasing by 2°C to 61°C per cycle and 61°C thereafter; extension, 72°C for 2 min (cycles 1–10), then 4 min (cycles 11–20) and finally 6 min (cycles 21–30). In the primary PCR reaction 1 μl of ligation product was amplified in 50 μl using 400 ng WT1 primer (5′-CCA AGG GCC GTG AGG ATA GCG GAA G-3′) and 250 ng splinkerette/vectorette primer (5′-CGA ATC GTA ACC GTT CGT ACG AGA A-3′). Secondary PCR was performed using 1 μl primary PCR product, 400 ng internal WT1 primer (5′-GCA CGC AGG CAC TGG CCC CCG ACA T-3′) and 250 ng internal splinkerette/vectorette primer (5′-TCG TAC GAG AAT CGC TGT CCT CTC C-3′). Tracks 1 and 2 show 10 μl vectorette and splinkerette secondary PCR product respectively resolved on a 1% agarose gel. The marker is the 1 kb ladder (Gibco-BRL). No product was visible after primary PCR for either the vectorette or the splinkerette. Tracks 3 and 4 show results of Southern blot analysis of tracks 1 and 2, respectively, with a 1.1 kb probe derived from within the target PCR product.

REFERENCES

1 Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J.C. and Markham, A.F. (1990) *Nucleic Acids Res.* **18**, 2887–2890.
2 Arnold, C. and Hodgson, I.J. (1991) *PCR Methods Appl.* **1**, 39–42.
3 Tadokoro, K., Oki, N., Fujii, H., Oshima, A., Inoue, T. and Yamada, M. (1992) *Jpn. J. Cancer Res.* **83**, 1198–1203.
4 Sommer, R. and Tautz, D. (1989) *Nucleic Acids Res.* **17**, 6749.
5 Evans, K.L., Devon, R.S., Brookes, A.J., Porteous, D.J. *et al.*, (1995) *Genomics*, in press.
6 Chou, Q., Russell, M., Birch, D.E., Raymond, J. and Bloch, W. (1992) *Nucleic Acids Res.* **20**, 1717–1723.
7 Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. and Mattick, J.S. (1991) *Nucleic Acids Res.* **19**, 4008.

GENOMICS 28, 420–428 (1995)

A Contiguous Clone Map over 3 Mb on the Long Arm of Chromosome 11 across a Balanced Translocation Associated with Schizophrenia

KATHRYN L. EVANS,* JOHN BROWN,*,† YOSHIRO SHIBASAKI,* REBECCA S. DEVON,* LIN HE,* BENOIT ARVEILER,*,‡ SHEILA CHRISTIE,* JOHN C. MAULE,* DAVID BAILLIE,* EUAN M. SLORACH,* SUSAN M. ANDERSON,* JOHN R. GOSDEN,* JOELLE PETIT,‡ ANDREAS WEITH,§ CHRISTINE M. GOSDEN,*,¶ DOUGLAS H. R. BLACKWOOD,*,† DAVID M. ST. CLAIR,*,† WALTER J. MUIR,*,† ANTHONY J. BROOKES,* AND DAVID J. PORTEOUS*,[1]

*MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, United Kingdom; †Department of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, United Kingdom; ‡Departement de Biochimie Medicale et Biologie Moleculaire, University of Bordeaux II, 146 Rue Leo Saignat, 33076 Bordeaux Cedex, France; §Research Institute of Molecular Pathology, Dr Bohr-Gasse 7, A-1030 Vienna, Austria; and ¶Department of Obstetrics and Gynaecology, Royal Liverpool University Hospital, Prescott Street, Liverpool L7 8XP, United Kingdom

Received October 20, 1994; accepted March 14, 1995

Forty-nine clones derived by microdissection of a schizophrenia-associated t(1;11)(q42.1;q14.3) breakpoint region have been assigned by somatic cell hybrid mapping to seven discrete intervals on the long arm of human chromosome 11. Eleven of the clones were shown to map to a small region immediately distal to the translocation breakpoint on 11q. A 3-Mb contiguous clone map of this region was established by isolation of corresponding YAC recombinants. The contig was oriented and shown to traverse the translocation breakpoint by FISH and microsatellite marker analysis. This contig will facilitate the isolation of candidate sequences whose expression may be affected by the translocation. © 1995 Academic Press, Inc.

INTRODUCTION

Progress toward a comprehensive map of the human genome has been greatly accelerated by the advent of YAC cloning for physical mapping and the utilization of microsatellite repeats for genetic mapping. Indeed, framework physical and genetic maps of the entire genome are now in place (NIH/CEPH Collaborative Mapping Group, 1992; Weissenbach *et al.,* 1992; Cohen *et al.,* 1994). Nevertheless, the limited resolving power of conventional linkage mapping and the effect of YAC chimerism on contig assembly mean that considerable effort is still required to map accurately subchromosomal regions of particular clinical interest. Human chromosome 11 is one of the better mapped chromosomes (Evans, 1993; James *et al.,* 1994), but as is also true for all other chromosomes, there is a relative dearth of markers in certain subregions. This reflects, at least in part, the manner in which markers are selected and the inherent biases in certain cloning procedures for or against subclasses of nonrandomly distributed DNA sequences. One such poorly mapped region is 11q14–q21, between tyrosinase (TYR) and the marker D11S388 on the long arm of human chromosome 11, a distance estimated to be ~8 cM (NIH/CEPH Collaborative Mapping Group, 1992). We have previously shown that a balanced translocation, t(1;11)(q42.1;q14.3), which cosegregates with schizophrenia and other severe mental illness in a large Scottish pedigree (St. Clair *et al.,* 1990), breaks in this region (Fletcher *et al.,* 1993).

To characterize the molecular nature of this translocation breakpoint, it was first necessary to obtain new markers for this region. To this end, we undertook to microdissect and microclone the region directly from the der(1) translocation chromosome (Muir *et al.,* 1995). *In situ* hybridization "painting" of the microdissection library back onto metaphase chromosomes from the translocation patient had shown that approximately 50% of the cloned material derived from chromosome 1, immediately proximal to the breakpoint, and approximately 50% from chromosome 11, immediately distal to the breakpoint. The cloned inserts averaged 1.5 kb in length and thus were potentially suitable for direct hybridization screening of genomic libraries and Southern blots. We have also described previously a panel of somatic cell hybrids bearing reduced human chromosomes 11, which permit high-resolution local-

[1] To whom correspondence should be addressed. Telephone: (44) 31-332 2471, x2121. Fax: (44) 31-343 2620. Email: davep@hgu.mrc.ac.uk.

0888-7543/95 $12.00

ization of such markers on the long arm (Fletcher *et al.*, 1993; Evans *et al.*, 1993, 1995; Slorach *et al.*, 1995).

The objectives of this study were (a) to localize regionally microdissection-derived clones (MDs) on the long arm of chromosome 11 by somatic cell hybrid mapping, (b) to use those mapping closest to the translocation breakpoint as probes to isolate YAC recombinants, and (c) to construct a contiguous clone map across the region of chromosome 11 disrupted by the schizophrenia-associated translocation. This would provide a starting point for studies aimed at determining the molecular nature of that breakpoint and the effect it may have on the expression of neighboring genes. This study demonstrates the utility of the microdissection clones as hybridization probes. Using just five of the microdissection clones that map to the interval on 11q closest to the breakpoint, we were able to establish a contig map of ~3 Mb with only a single gap. Isolation of critical end clones and one further round of YAC library screening allowed that gap to be filled and validated the contig. Fluorescence *in situ* hybridization and microsatellite marker analysis were used to establish orientation and to confirm that this contiguous clone map on chromosome 11 did extend over the translocation breakpoint.

MATERIALS AND METHODS

Somatic cell hybrid lines. A subset of hybrid cell lines from our previously described chromosome 11q mapping panel (Fletcher *et al.*, 1993; Evans *et al.*, 1993, 1995; Slorach *et al.*, 1995) was used to localize the microdissection clones regionally. The hybrids MIS 7.4, PG48, CF52, WJX 5.4, WJX 11.2, and E67.4 constitute the minimum set that defines all discernible intervals distal to the t(1;11)(q42.1;q14.3) translocation breakpoint and are described briefly below. MIS 39.8 was used to define the region on human chromosome 11 centromeric to the t(1;11)(q42.1;q14.3) translocation breakpoint.

(a) MIS 7.4 and MIS 39.8 are mouse hybrids formed by fusion of a human lymphoblastoid cell line carrying the t(1;11)(q42.1;q14.3) and retain several human chromosomes but carry only the der(1) and der(11) translocation chromosomes, respectively, in the absence of any normal chromosome 1 or 11 material.

(b) PG48 is a hamster hybrid that retains several human chromosomes but whose only human chromosome 11 material is a translocated chromosome 11, which lacks 11q22–qter.

(c) CF52 is a mouse hybrid that retains several human chromosomes but has as its only human chromosome 11 component a t(11;16)(q13;p11) translocation chromosome.

(d) The X irradiation hybrids, WJX 5.4 and WJX 7.4, are hamster hybrids that contain multiple subfragments of human chromosome 11 (derived from the human chromosome 11-only hamster hybrid J1), obtained by positive selection for expression of human-specific cell surface markers encoded by genes on 11q.

(e) The HRAS1-selected chromosome-mediated gene transformant, E67.4, is a mouse hybrid that carries a single human transgenome comprising only human chromosome 11 material, derived from both the short and the long arms with multiple interstitial deletions.

Cell culture methods and DNA isolation. Standard conditions were used, as described previously (Fletcher *et al.*, 1993).

Mapping of microdissection clones to somatic cell hybrids. The microdissection, microcloning, and preparation of inserts is described in detail elsewhere (Muir *et al.*, 1995). Briefly, inserts of individual microdissection clones were amplified by PCR using primers flanking the *Eco*RI cloning site. Vector sequences were removed by *Eco*RI digestion and electrophoresis through 1% LMP agarose. The inserts were labeled by random priming and then hybridized directly to Southern blots of *Eco*RI restriction digested hybrid DNA (~5 μg/track). Southern blots were hybridized at 68°C and washed to 0.1× SSC at 68°C. Hybridization signal was detected by a Molecular Dynamics PhosphorImager after exposure for 24 h.

Mapping of D11S931 and D11S873 to somatic cell hybrids and YAC recombinants. Previously published primers and PCR conditions for amplification of D11S931 (Weissenbach *et al.*, 1992) and D11S873 (Litt *et al.*, 1993) were used.

YAC libraries. The ICI YAC library (Anand *et al.*, 1990) was gridded by one of us (B.A.) at the UK HGMP Resource Centre, London. The library consists of approximately 35,000 YAC recombinants gridded on 23 filters of 8 × 12 cm using a 96-pin device. The average insert size is approximately 350 kb, and the theoretical complexity of the library is therefore >3.5 genome equivalents. Duplicate filters of three of the large-insert high-density gridded ICRF YAC libraries (Lehrach *et al.*, 1990; Larin *et al.*, 1991) comprising ~21,000 clones in total and representing ~3 genome equivalents were made available (courtesy of Hans Lehrach) through the ICRF Reference Library Database.

YAC library screening. Microdissection clones from the interval immediately distal to the translocation breakpoint on chromosome 11 were prepared as described above and used to screen the ICI and ICRF YAC libraries by filter hybridization. The microdissection clones chosen, MD 283 (ICI library screening only), MD 122, MD 176, MD 220, and MD 471, were those that failed to hybridize to human Cot I DNA and that showed no evidence of repetitive content, as judged by hybridization to Southern blots of human genomic DNA digests.

Isolation of YAC DNA. DNA was prepared from YAC recombinants by standard procedures described elsewhere (Arveiler, 1994).

Isolation of YAC end clones. The vectorette method used was as originally described (Riley *et al.*, 1990) using *Sau*3AI compatible ends. Total yeast DNA was digested with *Bam*HI, *Bgl*II or *Bcl*I before ligation. The PCR employed a "hot-start" (Chou *et al.*, 1992). Conditions were 95°C for 40 s, 72°C for 4 min (cycles 1 to 10), 5 min (cycles 11 to 20), 7 min (cycles 21 to 30), and finally 10 min (cycle 31).

The splinkerette method was essentially as described elsewhere (Devon *et al.*, 1995). Total yeast DNA was digested with *Bam*HI plus *Bgl*II and ligated to duplexed oligonucleotides with *Sau*3AI compatible ends. The PCR employed hot-start (Chou *et al.*, 1992) and "touch down" (Don *et al.*, 1991) conditions. Secondary amplification with a primer internal to the left arm YAC primer was sometimes used. The YAC arm primers were derived from those described previously for inverse PCR (Arveiler and Porteous, 1991). Splinkerette PCR: left arm YAC primer, 5′ CCC GTC CTG TGG ATC AAT TC 3′; left arm internal, 5′ GCC AAG TTG GTT TAA GGC GC 3′; right arm YAC primers, 5′ GCT CCT TCC TTC GTT CTT CC 3′ or 5′ CCA CCA TAC CCA CGC CGA AAC AA 3′. Splinkerette oligonucleotides: upper, 5′ CGA ATC GTA ACC GTT CGT ACG AGA GTT CGT ACG AGA ATC GCT GTC CTC TCC AAC GAG CCA AGG 3′; lower, 5′ GAT CCC TTG GCT CGT TTT TTT TTG CAA AAA 3′. Splinkerette primer, 5′ CGA ATC GTA ACC GTT CGT ACG AGA A 3′; PCR conditions, denaturation, 94°C, 30 s (cycle 1) and then 15 s to end; annealing, 70 → 60°C (by 2°C per cycle) and then 60°C to end for 1 min; extension, initially 3 min (cycles 1 to 10), then 5 min (cycles 11 to 20), and finally 8 min (cycles 21 to 30) at 72°C.

Catch linkering of YAC recombinants. The method has been described in detail elsewhere (Shibasaki *et al.*, 1995). Briefly, YAC recombinants were separated from the yeast chromosomes by analytical PFGE (~5 μg of total yeast DNA), digested with *Sau*3AI, ligated to a compatible oligonucleotide duplex (upper strand, 5′ CCG AAT TCT AGA GTC GACC 3′; lower strand, 5′ GAT CGG TCG ACT CTA GAA TTC G 3′), and PCR primed and amplified using the upper-strand oligonucleotide as follows: 35 cycles of 93°C for 15 s; 54°C for 30 s; 72°C for 1 min (cycles 1–10), then 2 min (cycles 11–20), and finally 4 min (cycles 21–35).

TABLE 1

Regional Mapping of Microdissection Clones

MD clone	GDB No.	MIS 7.4	CF52	PG48	WJX 7.4	WJX 5.4	E67.4
MD 543	D11S2107	+	+	+	+	−	+
MD 104	D11S2108	+	+	+	+	−	+
MD 282	D11S2109	+	+	+	+	−	+
MD 437	D11S2110	+	+	+	+	−	+
MD 220	D11S2111	+	+	+	+	−	+
MD 286	D11S2112	+	+	+	+	−	+
MD 283	D11S2113	+	+	+	+	−	+
MD 176	D11S2114	+	+	+	+	−	+
MD 122	D11S2115	+	+	+	+	−	+
MD 412	D11S2116	+	+	+	+	−	+
MD 471	D11S2117	+	+	+	+	−	+
	D11S873	+	+	+	+	−	+
MD 174	D11S2129	+	+	+	−	−	−
MD 139	D11S2130	+	+	+	−	−	−
	D11S388	+	+	+	−	−	−
MD 128	D11S2118	+	+	+	−	−	+
MD 118	D11S2119	+	+	+	−	−	+
MD 310	D11S2120	+	+	+	−	−	+
MD 130	D11S2122	+	+	+	−	−	+
MD 040	D11S2123	+	+	+	−	−	+
MD 159	D11S2124	+	+	+	−	−	+
MD 129	D11S2125	+	+	+	−	−	+
MD 377	D11S2126	+	+	+	−	−	+
MD 421	D11S2127	+	+	+	−	−	+
MD 281	D11S2128	+	+	+	−	−	+
MD 074	D11S2121	+	+	+	−	−	+
MD 558	D11S2131	+	+	+	−	+	+
MD 330	D11S2132	+	+	+	−	+	+
MD 173	D11S2138	+	+	+	−	+	+
MD 038	D11S2139	+	+	+	−	+	+
MD 021	D11S2140	+	+	+	−	+	+
MD 063	D11S2141	+	+	−	−	+	+
MD 136	D11S2142	+	+	−	−	+	+
MD 027	D11S2143	+	+	−	−	+	+
MD 151	D11S2144	+	+	−	−	+	+
MD 509	D11S2145	+	+	−	−	+	+
MD 211	D11S2146	+	+	−	−	+	+
MD 116	D11S2147	+	+	−	−	+	+
MD 388	D11S2148	+	+	−	−	+	+
MD 596	D11S2149	+	+	−	−	+	+
MD 075	D11S2150	+	+	−	−	+	+
MD 399	D11S2151	+	+	−	−	+	+
	CLG	+	+	−	−	+	+
	STMY1	+	+	−	−	+	+
	D11S385	+	+	−	−	+	+
MD 110	D11S2152	+	+	−	−	−	+
MD 453	D11S2153	+	+	−	−	−	+
MD 214	D11S2154	+	+	−	−	−	+
	NCAM	+	+	−	+	−	+
	DRD2	+	+	−	+	−	−
MD 096	D11S2155	+	+	−	−	+	−
MD 054	D11S2156	+	+	−	−	+	−
MD 162	D11S2157	+	+	−	−	+	−
MD 044	D11S2158	+	+	−	−	+	−
MD 215	D11S2159	+	+	−	−	+	−
MD 277	D11S2160	+	+	−	−	+	−
	D11S351	+	+	ND	−	+	−
	THY1	+	+	ND	−	+	−

Note. The table summarizes the hybridization results of 49 microdissection clones assigned to chromosome 11 and regionally mapped using the chromosome 11q mapping panel. Established markers mapped previously to the panel are included for reference. In many cases, microdissection clones were placed in a previously defined interval, but in several cases (e.g., MD 021; see Fig. 1b) the segregation pattern identified a new interval. In this situation markers were arranged in the order that assumed the least number of independent blocks of human DNA in the various hybrids. +, present; −, absent; ND, not determined.

TABLE 2

First-Generation YAC Contig Map

YAC designation	MD 543	MD 104	MD 282	MD 437	MD 220	MD 286	MD 283	MD 176	MD 122	MD 412	MD 471
ICRF y901 D0485	+	+	+	+	+						
ICI yac 35IF5			+	+	+	+	+				
ICI yac 6DB11					+	+	+				
ICI yac 14IA11					+	+	+				
ICRF y901 H0515					+						
ICRF y900 A0846								+			
ICRF y900 D11155								+	+	+	
ICI yac 11EA9								+			
ICRF y900 H07131								+			
ICI yac 15IE11								+			
ICI yac 16GB9								+	+		
ICRF y900 C04151									+		
ICI yac 11BA9									+		
ICI yac 2AH11									+		
ICI yac 33EB11									+		
ICRF y900 A04135										+	+
ICI yac 8AB2										+	+
ICRF y900 D0165											+
ICRF y905 H0921											+
ICI yac 34GE5											+
ICI yac 38CE7											+
ICI yac 5FD7											+

Note. The table summarizes the results of hybridization of each of the 11 microdissection clones in the interval closest to the translocation breakpoint to restriction-digested DNA from all 23 of the YACs obtained by screening with MDs 220, 283, 176, 122, and 471.

The whole catch linker product was directly labeled for use as a FISH probe by PCR amplification using the upper-strand primer biotinylated at the 5′ end (Shibasaki *et al.,* 1995). Catch linkered products were also cloned by *Sau*3AI digestion and ligation into the *Bam*HI cloning site of pBluescribe (Stratagene). Inserts from individual colonies were recovered by reamplification using universal primers flanking the cloning site (Gussow and Clackson, 1989).

Fluorescence in situ hybridization. Interphase and prometaphase chromosomes were prepared essentially as described previously (Shibasaki, 1994) from fresh peripheral blood lymphocytes or lymphoblastoid cell lines from individuals either with a normal karyotype or bearing the t(1;11)(q42.1;q14.3) chromosomes. Cosmids corresponding to selected MD clones were obtained by hybridization to gridded filters of the ICRF Reference Library No. 107(L4/FS11) constructed by D. Nizetic (Lehrach *et al.,* 1990). YAC *in situ* hybridization utilized catch linker material prepared as described above and in detail elsewhere (Shibasaki *et al.,* 1995). Chromosomal *in situ* suppression hybridization was used throughout (Lichter *et al.,* 1988; Shibasaki, 1994). Cosmid DNA and *Alu*-PCR products of YACs were labeled by nick-translation using either bio-16–dUTP or DIG-11–dUTP. Catch-linkered YAC inserts were labeled directly by PCR using biotinylated primer (Shibasaki *et al.,* 1995). The signal detection for the biotin–avidin system was performed according to Pinkel *et al.* (1986) using FITC–avidin DCS (Vector Labs) and biotinylated anti-avidin antibody (Vector Labs). For multicolor FISH, DIG-labeled probes were visualized together with biotin-labeled probes by using mouse monoclonal anti-DIG antibody (Boehringer Mannheim), rabbit anti-mouse antibody conjugated to Texas red (Jackson ImmunoResearch Lab.), and mouse anti-rabbit antibody conjugated to Texas red (Jackson ImmunoResearch Lab.). To generate yellow color, equal amounts of biotin- and DIG-labeled probes were cohybridized and detected by FITC and Texas red. FISH slides were examined under a Zeiss AxioPhoto microscope equipped with a Photometrics CCD

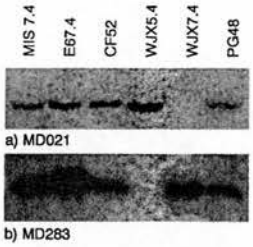


FIG. 1. Regional mapping of microdissection clones to the hybrid panel. The figure shows representative results with two microdissection clones, (**a**) MD 021 and (**b**) MD 283. Both clones hybridize to MIS 7.4, the hybrid representing the microdissected chromosome. The pattern of hybridization observed in the other hybrids was used to assign each clone to an interval distal to the t(1;11)(q42.1;q14.3) translocation breakpoint (see Table 1). The marker is the 1-kb ladder (Gibco BRL). The size of the hybridizing fragment corresponds to that of the particular insert, as *Eco*RI was used both to produce the microdissection clone insert and to digest the target genomic DNA.

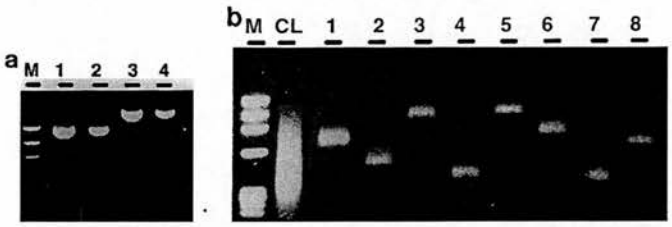


FIG. 2. Isolation of YAC end clones and "catch linker" products. (**a**) The molecular weight marker is *Hae*III-digested ϕX174. Lanes **1** to **4** show end clones isolated by splinkerette PCR from the left arm of ICI yacs 19EE2, 21BG4, 2AH11, and 33EB11, respectively. Note that the products from 19EE2 and 21BG4 are indistinguishable in size, part of the evidence for these YACs being an identical set, together with ICI yac 11BA9. (**b**) The molecular weight marker is *Hae*III-digested ϕX174. CL is the primary catch linker PCR product. Lanes **1** to **8** show amplified inserts after cloning into pBluescribe, as described under Materials and Methods.

TABLE 3

Localization and Characterization of MD-Derived Cosmid and YAC Clones by PFGE and FISH

Clone designation	Size	Location	Note	End clones mapped to
cos Tyr 7H3 (TYR)	<40 kb	11q14.3		
ICI yac 12GE1 (TYR)	350 kb	11q14.3	CL	
ICRF c107 A0551 (MD104)	<50 kb	11q14.3–q21		
ICRF c107 F790 (MD220)	<50 kb	11q14.3–q21		
ICRF c107 A0543 (MD283)	<50 kb	11q14.3–q21		
ICRF y901 D0485	1.3 Mb	11q14.3–q21	CL, cosmid subclones (cYS4)	L: 36IF5, 6DB11, 14IA11
ICI yac 36IF5	600 kb	11q14.3–q21, 5p14	*Alu*-PCR, CL, chimeric, cosmid subclones (cYS2)	L: D0485 R: chimeric end
ICI yac 14IA11	350 kb	11q14.3–q21	CL	
ICRF y900 A0846	600 kb	11q14.3–q21	CL	L: 36IF5, 6DB11, 14IA11 R: D11155, 11EA9, 15IE11, 16GB9
ICRF y900 D11155	700 kb	11q14.3–q21	CL	L: A04135, 8AB2, D0165 R: A0846
ICI yac 11EA9	270 kb	11q14.3–q21	CL	L: D11155, 15IE11, 16GB9, 11BA9, C04151 R: A0846, 37GE11, 35IA2, D11155
ICI yac 15IE11	240 kb	11q14.3–q21	CL	
ICI yac 16GB9	330 kb	11q14.3–q21	CL, cosmid subclones (cYS3)	L: 11BA9, 2AH11 R: 11EA9, 15IE11
ICI yac 11BA9	300 kb	11q14.3–q21	*Alu*-PCR	L: 11EA9, 15IE11, 16GB9
ICI yac 2AH11	200 kb	11q14.3–q21	*Alu*-PCR	L: 15IE11, 16GB9, 11BA9,
ICI yac 33EB11	350 kb	11q14.3–q21, 18q23	*Alu*-PCR, CL, chimeric	L: chimeric end R: 16GB9, 11BA9, 2AH11
ICRF y900 A04135	700 kb	11q14.3–q21	CL	L: D11155, 11BA9, 2AH11, C04151 R: H0921
ICI yac 8AB2	260 kb	11q14.3–q21	*Alu*-PCR	R: A04135, 5FD7, D0165, H0921, 34GE5
ICI yac 38CE7	280 kb	11q14.3–q21, 19?	*Alu*-PCR, chimeric?	R: 8AB2, 34GE5
ICI yac 5FD7	250 kb	11q14.3–q21	CL	L: 8AB2, 34GE5, 38CE7
cosCJ52.4 (D11S388)	<50 kb	11q21		

Note. The table summarizes the cosmid and YAC clones used in FISH analysis, the insert sizes of the YACs as estimated by pulsed-field gel electrophoresis, their localization on the long arm of chromosome 11, and how they were labeled and identifies those shown or suspected to be chimeric. YAC clones from which end clones and subfragments were isolated are also identified. CL, catch linker plus PCR labeling; L, left arm YAC end clone; R, right arm YAC end clone.

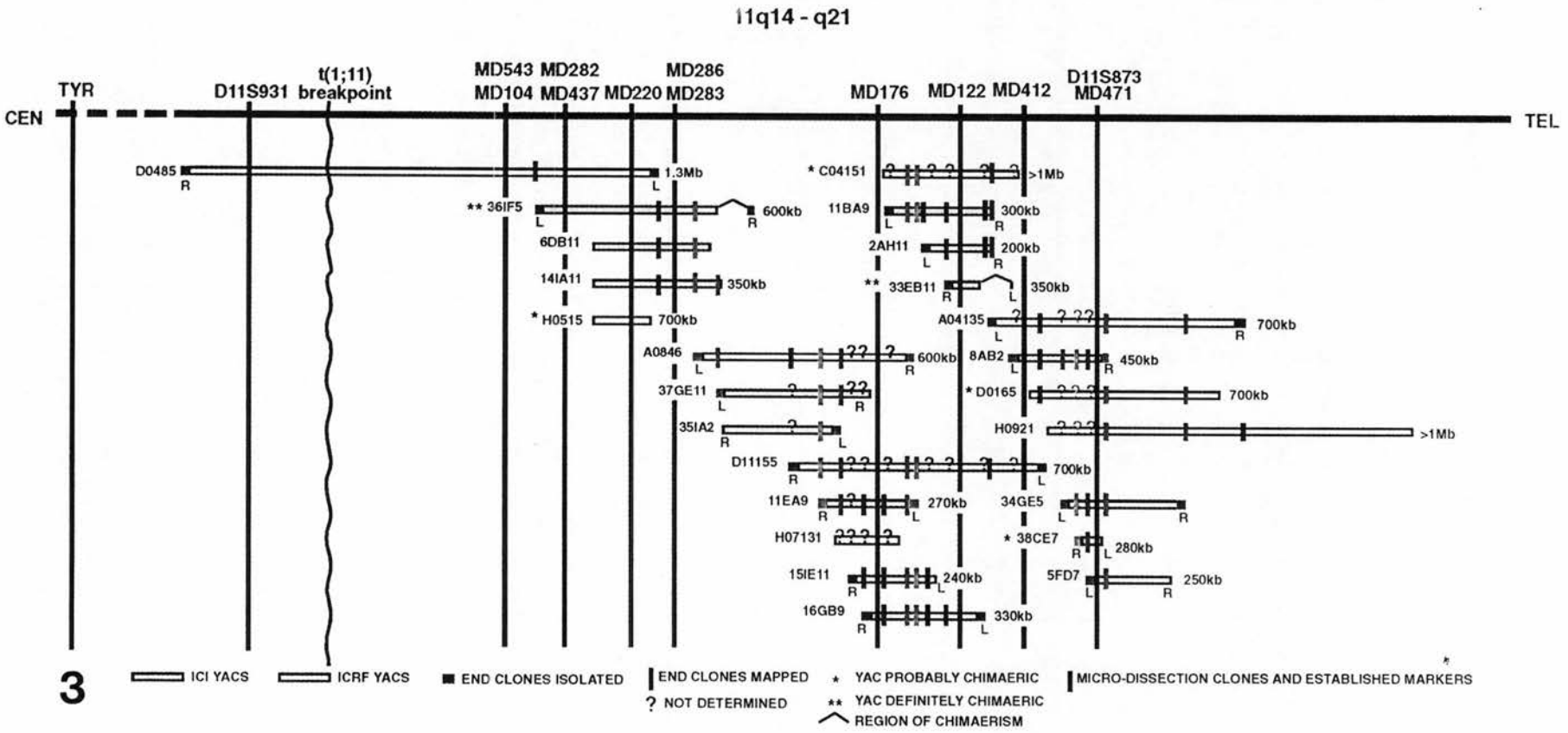
camera and an IPLab image analyzing system on an Apple Macintosh computer or a Leiz microscope equipped with a Bio-Rad MRC-600 confocal laser scanning system.

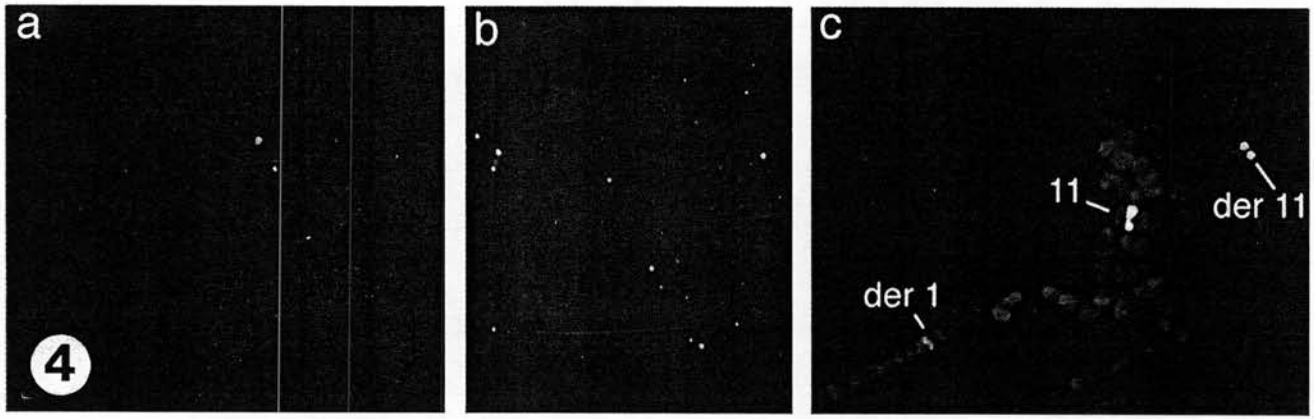
RESULTS

We have described elsewhere a somatic cell hybrid mapping panel that allows precise localization of markers on the long arm of human chromosome 11 (Fletcher *et al.,* 1993; Evans *et al.,* 1993, 1995; Slorach *et al.,* 1995). By combining somatic cell hybrids bearing translocation chromosomes with stable, positively selected radiation-reduced or chromosome-mediated gene transformant hybrids, it was possible to assemble a panel that facilitated high-resolution mapping in the region immediately distal to the t(1;11) breakpoint from a minimum of genomic DNAs. This has obvious advantages over the use of large sets of transient radiation hybrids (James *et al.,* 1994), not least the fact that these hybrids represent a stable replenishable mapping resource and allow direct hybridization to be used in addition to PCR-based mapping procedures. Eighty-six microdissection-derived clones (MDs) were mapped onto a subpanel of just six somatic cell hybrids that

FIG. 3. An 3-Mb contig on chromosome 11q14–q21. The figure shows the contig in diagrammatic form and approximately to scale. All YACs have been tested with all MDs, and their presence is indicated by a solid vertical bar. Isolated YAC end clones are indicated, and their presence in other YACs is marked by a bar. Not all of the YACs have been tested with all of the end clones, and this is indicated by a "?". Chimeric YACs are asterisked, and the region of chimerism is indicated where known. Only minimum addresses are shown for each YAC (see text and Tables 2 and 3 for full details).

FIG. 4. Orientation of the contig and breakpoint analysis by multicolor FISH. (**a**) Triple-color FISH analysis reveals the order of MD 104 (ICRF c107 A0551, red) and MD 220 (ICRF c107 F790, yellow) relative to TYR (ICI yac 12GE1, green), establishing the orientation of the contig. Note that the distance between TYR (green) and MD 104 (red) is much greater than that between MD 104 (red) and MD 220 (yellow). (**b**) The result of triple-color FISH using MD 283 (ICRF c107 A0543, green), ICRF y900 A0846 (red), and MD 176/122 (cYS3-pool, yellow), placing the YAC into the gap of the first-generation contig. This result also confirms the orientation of the contig with respect to the translocation breakpoint. (**c**) A partial metaphase plate of a patient carrying the reciprocal t(1;11) chromosomes hybridized with catch-linkered ICRF y901 D0485. Confined signals are seen on the der(11) as well as the der(1) and normal chromosome 11, demonstrating that the YAC crosses the translocation breakpoint.


11q14 - q21
TYR
CEN
TEL
D11S931
t(1;11) breakpoint
MD543 MD282 MD286
MD104 MD437 MD220 MD283
MD176 MD122 MD412
D11S873 MD471
D0485
1.3Mb
** 36IF5
600kb
6DB11
14IA11
350kb
* H0515
700kb
* C04151
>1Mb
11BA9
300kb
2AH11
200kb
** 33EB11
350kb
A04135
700kb
A0846
600kb
8AB2
450kb
37GE11
* D0165
700kb
35IA2
H0921
>1Mb
D11155
700kb
11EA9
270kb
34GE5
H07131
* 38CE7
280kb
15IE11
240kb
5FD7
250kb
16GB9
330kb
3
ICI YACS
ICRF YACS
END CLONES ISOLATED
END CLONES MAPPED
? NOT DETERMINED
* YAC PROBABLY CHIMAERIC
** YAC DEFINITELY CHIMAERIC
REGION OF CHIMAERISM
MICRO-DISSECTION CLONES AND ESTABLISHED MARKERS


a
4
b
c
11
der 11
der 1

included the hybrid, MIS 7.4, bearing the schizophrenia-associated der(1) translocation chromosome. Representative hybridization results are shown in Fig. 1. Of these 86 MDs, 49 were localized to intervals on 11q, while the remaining 37 (43%) hybridized only to MIS 7.4 and were therefore provisionally assigned to chromosome 1.

All 49 of the chromosome 11 assigned microdissection clones mapped to or above the interval defined by THYI (see Table 1), in general agreement with the expectation from our previous FISH "painting" analysis of the library (Muir *et al.,* 1995). Microdissection clones were mapped to all intervening intervals defined by the hybrid panel, with the exception of those marked by NCAM and by DRD2. Other studies have shown that these genes map within ~500 kb of each other (McConville *et al.,* 1990). The microdissection clones also define two new intervals on chromosome 11. Most importantly, 11 map to a region immediately distal to the breakpoint, as defined by the microsatellite D11S873, a recently defined marker that lies centromeric to D11S388 (Litt *et al.,* 1993; Evans *et al.,* 1995). These results are therefore in keeping with our previous estimate that the microdissected region on chromosome 11 corresponds to ~15 Mb (Muir *et al.,* 1995) and the expectation that the markers approximate to a normal distribution around the translocation breakpoint itself. The distance separating TYR and D11S388 has been estimated to be ~8 cM and that between D11S873 and D11S388 to be ~2.4 cM (NIH/CEPH Collaborative Mapping Group, 1992; Weissenbach *et al.,* 1992). It was therefore considered likely that the 11 MDs mapped to the "breakpoint–D11S873" interval would be sufficient to isolate a complete YAC contig of the region.

Isolation of YAC Recombinants

To construct a contiguous clone map toward the translocation breakpoint, five repeat-free microclones (MD 220, MD 283, MD 176, MD 122, and MD 471) were selected from this set of 11 and used as direct hybridization probes to gridded filters of YAC recombinants. Two libraries were screened, the ICI YAC library (Anand *et al.,* 1990) (courtesy of the UK HGMP Resource Centre) (with all five MDs) and the ICRF Reference Library (Larin *et al.,* 1991) (courtesy of Dr. Hans Lehrach) (MD 283 not used). A total of 23 independent YAC recombinants were obtained. These were reprobed with the remaining six microdissection clones to assemble a first-generation contig map as shown in Table 2. By this simple screening procedure, a contig map was assembled with just a single gap (between MD 283 and MD 176), but it contained no information regarding orientation (relative or absolute), net size, or extent of overlap between YACs.

Isolation of YAC Endclones and Gap Closure

To confirm the first-generation contig map and achieve closure of the single gap, end clones were isolated from selected YAC recombinants by either the vectorette (Riley *et al.,* 1990) or the splinkerette (Devon *et al.,* 1995) methods (Fig. 2a) and probed onto the primary set of 23 YACs. In addition, selected YAC recombinants were purified and ligated to catch linkers for PCR amplification (Fig. 2b). Single-copy subclones isolated from plasmid libraries, derived from catch linkered material, were also used to confirm overlaps between YAC recombinants and as additional markers for genomic hybridization. YAC recombinants were sized by pulsed-field gel electrophoresis as summarized in Table 3 and analyzed with end clones. Secondary screening of the ICI YAC library was carried out with the right arm end clone of ICI yac 11EA9, which identified ICI yacs 37GE11 and 35IA2. Isolation of a left arm end clone from ICI yac 37GE11 confirmed the link from ICRF y900 A0846 through the ICI yacs 14IA11 and 36IF5 and therefore to ICRF y900 D0485 (see Table 3 and Fig. 3). This analysis allowed us to establish the consensus map shown in Fig. 3. Note that the marker positions are relative, not absolute. The estimated end-to-end length of the contig map is ~3.5 Mb. When the most telomeric and only partially characterized YAC ICRF y905 H0921 is excluded, the validated contig, which extends across the translocation breakpoint, is itself greater than 2.8 Mb. This follows from the insert sizes of ICRF y901 D0485, ICRF y900 A0846, ICI yac 2AH11, and ICRF y900 A04135, which alone sum to ~2.8 Mb, but this does not take into account the two internal regions represented by (a) MD 286 and MD 283 (three positive overlapping YACs) or (b) the left end clone of ICI yac 11EA9 (five positive overlapping YACs). The core contig, that is, excluding the terminal regions represented only by D0485 and H0921, respectively, sums to ~2.2 Mb. This agrees well with the size estimate based on (i) our previous mapping results, (ii) the known genetic distance between flanking markers, and (iii) the estimated physical distance based on our previous FISH analysis (Muir *et al.,* 1995).

Establishing Orientation and Crossing the Translocation Breakpoint

To establish the orientation of the cloned contig, FISH analysis on prometaphase and interphase chromosomes was undertaken, with reference to established, previously mapped markers (Fig. 4). Figure 4a shows a representative result, using the TYR locus (in the form of ICI yac 12GE1) as a reference marker in green and cosmids isolated using two microdissection clones MD 104 and MD 220 in red and yellow, respectively. The demonstration that MD 104 maps between TYR and MD 220 establishes orientation of the contig with respect to the chromosome 11 centromere and therefore the translocation breakpoint. This result was confirmed by independent analysis of (a) additional cosmid clones (either derived by screening with microdissection clones or subcloned from YAC recombinants) and (b) catch linkered total YAC probes (Table 3). Fig-

ure 4b shows a similar analysis with cosmid clones corresponding to MD 283 (green) and MD 176/122 (yellow) plus catch linkered ICRF y900 A0846 (red).

The orientation of the contig as determined by FISH predicts that the YAC recombinant ICRF y901 D0485 maps closest to the translocation breakpoint. To test this, the YAC insert was purified by catch linker amplification and hybridized to lymphocyte metaphase spreads from a schizophrenic patient carrying the translocation (Fig. 4c). Signal was seen on both the der(1) and the der(11) chromosomes, establishing that this YAC recombinant does cross the translocation breakpoint. In addition, the signal was confined to the normal chromosome 11 and to the derived chromosomes 1 and 11, confirming that this YAC recombinant is nonchimeric. Independent confirmation of this result was obtained by PCR analysis of somatic cell hybrid DNAs and the YAC recombinant ICRF y901 D0485 using primers specific for the microsatellite repeat marker D11S931 (Weissenbach *et al.,* 1992). Two fragments of 261 and 265 bp were amplified from total genomic DNA from a t(1;11) subject. No band was amplified from MIS 7.4, the patient-derived hybrid that contains the der(1) chromosome as its sole chromosome 11 material. Only one band, of 265 bp, was amplified from MIS 39.8, the patient-derived hybrid that carries the der(11) chromosome as its sole chromosome 11 material. PCR analysis of the YAC recombinants showed that only ICRF y901 D0485 was positive for D11S931. These results demonstrate that the t(1;11) breakpoint lies within the region flanked by D11S931 (centromeric) and MD 104 (telomeric) and represented by ICRF y901 D0485.

DISCUSSION

This study demonstrates the value of microdissection for establishing a very high density of region-specific markers. Forty-nine new markers have been regionally localized on the long arm of human chromosome 11, of which more than a fifth map to a small interval immediately telomeric to the t(1;11) breakpoint and previously identified by only a single marker. We have shown the utility of these markers for direct hybridization mapping to somatic cell hybrids and for screening of gridded YAC libraries. The value of panels of stable somatic cell and chromosome-reduced hybrids for accurate subregional mapping should also be emphasized. The six somatic cell hybrids together define a minimum of nine discrete intervals in the region 11q14.3–q23 and provide new and useful markers for seven of these. We have been able to establish an extensive contiguous clone array of YAC recombinants with a minimum number of screening steps. The method compares favorably with PCR-based methods of screening YAC superpools and pools for individual recombinants, a procedure prone to PCR amplification biases that may mean that valuable recombinants are not recovered.

In this regard, it is of interest to note that three examples of duplicate YAC recombinants were identified upon screening the ICI YAC library. Insert size, Cot I hybridization, fingerprint analysis, and hybridization with end clones and internal probes were used to establish identity (data not shown). Two sets are probably artifacts of the original gridding procedure (ICI yacs 37GE11 and 37GE12 are identical, as are ICI yacs 34GE5 and 34GF5). The third set comprising ICI yacs 11BA9, 19EE2, and 21BG4 is presumably genuine replica clones. These findings attest to the efficiency of the filter screening procedure.

FISH analysis with reference to known flanking markers and to the translocation breakpoint was initially used to establish the orientation of the cloned array (Fig. 4). This was subsequently confirmed by independent marker analysis on somatic cell hybrids. Although the microdissection clones themselves are too small to be used reliably for *in situ* hybridization, it is relatively trivial to obtain corresponding cosmids, again by direct filter hybridization. Furthermore, high-resolution FISH analysis is possible using minute quantities of YAC recombinants purified by analytical pulsed-field gel electrophoresis and subsequently amplified by catch linker-based PCR (Shibasaki *et al.,* 1995). As a by-product of this analysis, we were able to assess the degree of chimerism among the YAC recombinants. Three were shown to be definitely chimeric by FISH analysis and/or hybridization analysis with relevant end clones, microdissection clones, and/or catch linker-derived markers. A further three are likely to be chimeric based on provisional evidence from one or the other of the above analyses. We conclude that the level of chimerism is ~12.5% in both YAC libraries, but without any apparent bias toward those YACs with larger inserts.

This study was undertaken as part of our concerted effort to investigate the molecular basis of psychiatric disorders. While it remains to be seen whether the translocation described here is causally related to schizophrenia, the presence of a discrete breakpoint on a chromosome provides a precise target for positional cloning and an unique opportunity to investigate the molecular architecture of the translocation and its possible relationship to the manifestation of the disorder. The contiguous clone map described here establishes the key reagents necessary to achieve this aim. By applying these same methodological procedures, it should be equally possible to assemble contiguous clone maps rapidly for other intervals on chromosome 11 (and indeed chromosome 1) defined by regionally assigned microdissection clones.

ACKNOWLEDGMENTS

We thank the other members of the Psychiatric Genetics Group for additional help and information, Sarah Smith and Ramnath Elaswarapu (UK HGMP Resource Centre, Harrow, UK) for help in gridding the ICI YAC library and for providing YAC clones, Lorna Mitchell for preparing the manuscript, and Norman Davidson and colleagues for preparing the figures. W.M. is supported by an MRC

Clinical Scientist Fellowship. D.St.C. is a Wellcome Trust Senior Clinical Research Fellow in Clinical Science. This work was supported by the Medical Research Council, the UK HGMP, and the Wellcome Trust.

REFERENCES

Anand, R., Riley, J. H., Butler, R., Smith, J. C., and Markham, A. F. (1990). A 3.5 genome equivalent multi access YAC library: Construction, characterisation and storage. *Nucleic Acids Res.* **18:** 1951–1956.

Arveiler, B. (1994). Immortalised cell lines: Their creation and use in gene mapping. *In* "Chromosome Analysis Protocols" (J. R. Gosden, Ed.), Humana, Clifton, NJ.

Arveiler, B., and Porteous, D. J. (1991). Amplification of end fragments of YAC recombinants by inverse-polymerase chain reaction. *Technique* **3:** 24–28.

Chou, Q., Russell, M., Birch, D. E., Raymond, J., and Bloch, W. (1992). Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Res.* **20:** 1717–1723.

Cohen, D., Chumakov, I., and Weissenbach, J. (1994). A first generation physical map of the human genome. *Nature* **366:** 698–701.

Devon, R. S., Porteous, D. J., and Brookes, A. J. (1995). Splinkerettes—Improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Res.*, in press.

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., and Mattick, J. S. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19:** 4008.

Evans, G. A. (1993). Integrating maps of human chromosome 11. *Curr. Opin. Genet. Dev.* **3:** 418–424.

Evans, K. L., Fantes, J., Simpson, C., Arveiler, B., Muir, W., Fletcher, J., van Heyningen, V., Steel, K. P., Brown, S. D. M., St. Clair, D., and Porteous, D. J. (1993). Human olfactory marker protein maps close to tyrosinase and is a candidate gene for Usher syndrome type I. *Hum. Mol. Genet.* **2:** 115–118.

Evans, K. L., van Heyningen, V., and Porteous, D. J. (1995). Placement and refined mapping of established and new markers on human chromosome 11q using a small panel of somatic cell hybrids. *Eur. J. Hum. Genet.* **3:** 42–48.

Fletcher, J. M., Evans, K. L., Ballie, D., Byrd, P., Hanratty, D., Leach, S., Julier, C., Gosden, J. R., Muir, W., Porteous, D. J., St. Clair, D., and van Heyningen, V. (1993). Schizophrenia-associated chromosome 11q21 translocation: Identification of flanking markers and development of chromosome 11q fragment hybrids as cloning and mapping resources. *Am. J. Hum. Genet.* **52:** 478–490.

Gussow, D., and Clackson, T. (1989). Direct clone characterization from plaques and colonies by the polymerase chain reaction. *Nucleic Acids Res.* **17:** 4000.

James, M. R., Richard, C. W., Schott, J.-J., Yousry, C., Clark, K., Bell, J., Terwilliger, J. D., Hazan, J., Dubay, C., Vignal, A., Agrapart, M., Imai, T., Nakamura, Y., Polymeropoulos, M., Weissenbach, J., Cox, D. R., and Lathrop, G. M. (1994). A radiation hybrid map of 506 STS markers spanning human chromosome 11. *Nature Genet.* **8:** 70–76.

Larin, Z., Monaco, A. P., and Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci. USA* **88:** 4123–4127.

Lehrach, H., Drmanac, R., Hoheisel, J., Larin, Z., Lennon, G., Monaco, A. P., Nizetic, D., Zehetner, G., and Poustka, A.-M. (1990). "Hybridization Fingerprinting in Genome Mapping and Sequencing in Genome Analysis," Vol. I, "Genetic and Physical Mapping" (K. E. Davies and S. M. Tilghman, Eds.), pp. 39–81, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Lichter, P., Cremer, T., Borden, J., Manuelidis, L., and Ward, D. C. (1988). Delineation of individual human chromosomes in metaphase and interphase cells by *in situ* suppression hybridization using recombinant DNA libraries. *Hum. Genet.* **80:** 224–234.

Litt, M., Kramer, P., Hauge, X. Y., Weber, J. L., Wang, Z., Wilkie, P. J., Holt, M. S., Mishra, S., Donis-Keller, H., Warnich, L., Retief, A. E., Jones, C., and Weissenbach, J. (1993). A microsatellite-based index map of human chromosome 11. *Hum. Mol. Genet.* **2:** 909–913.

McConville, C. M., Formstone, C. J., Hernandez, D., Thick, J., and Taylor, A. M. R. (1990). Fine mapping of the chromosome 11q22–23 region using PFGE, linkage and haplotype analysis: Localization of the gene for ataxia telangiectasia to a 5 cM region flanked by NCAM/DRD2 and STMY/CJ52.75, 2.22. *Nucleic Acids Res.* **18:** 4335–4343.

Muir, W. J., Gosden, C. M., Brookes, A. J., Fantes, J., Evans, K. L., Maguire, S. M., Stevenson, B., Boyle, S., Blackwood, D. H. R., St. Clair, D. M., Porteous, D. J., and Weith, A. (1995). Direct microdissection and microcloning of a translocation breakpoint region t(1;11)(q42.2;q21) associated with schizophrenia. *Cytogenet. Cell Genet.*, in press.

NIH/CEPH Collaborative Mapping Group (1992). A comprehensive genetic linkage map of the human genome. *Science* **258:** 67–87.

Pinkel, D., Straume, T., and Gray, J. W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. USA* **83:** 2934–2938.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J. C., and Markham, A. F. (1990). A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18:** 2887–2890.

Shibasaki, Y. (1994). High resolution mapping of the MYCN protooncogene at human chromosome 2p24.3 by fluorescence *in situ* hybridisation. *Cytogenet. Cell Genet.* **66:** 75–76.

Shibasaki, Y., Maule, J. C., Devon, R. S., Slorach, E. M., Gosden, J. R., Porteous, D. J., and Brookes, A. J. (1995). Catch-linker + PCR labelling: A simple method to generate fluorescence *in situ* hybridisation (FISH) probes from yeast artificial chromosomes (YACs). *PCR Methods Applications* **4:** 209–211.

Slorach, E. M., Polymeropoulos, M., Evans, K. L., Seawright, A., Fletcher, J. M., Porteous, D. J., and Brookes, A. J. (1995). Regional localisation of 19 expressed sequence tags to chromosome 11 using PCR amplification of somatic cell hybrids. *Cytogenet. Cell Genet.*, in press.

St. Clair, D., Blackwood, D., Muir, W., Carothers, A., Walker, M., Spowart, G., Gosden, C., and Evans, H. J. (1990). Association within a family of a balanced translocation with major mental illness. *Lancet* **336:** 13–16.

van Heyningen, V. (1994). Immortalised cell lines: Their creation and use in gene mapping. *In* "Chromosome Analysis Protocols" (J. R. Gosden, Ed.), Humana, Clifton, NJ.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., and Lathrop, M. (1992). A second-generation linkage map of the human genome. *Nature* **359:** 794–801.

Catch-linker+PCR Labeling: A Simple Method to Generate Fluorescence In Situ Hybridization Probes from Yeast Artificial Chromosomes

Yoshiro Shibasaki, John C. Maule, Rebecca S. Devon, Euan M. Slorach, John R. Gosden, David J. Porteous, and Anthony J. Brookes

Medical Research Council (MRC) Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, Scotland, UK

A simple and efficient method to generate hapten-labeled DNA fragments from a trace amount of YAC DNA isolated by PFGE is described. After agarase digestion of the gel slice containing the resolved YAC recombinant, the purified DNA is digested with *Sau*3AI and a compatible CL oligonucleotide duplex ligated on. A probe is generated by PCR amplification using a primer complementary to the CL with a single biotin moiety incorporated at the 5′ end. When used as a FISH probe, this material yields mapping results superior to *Alu*–PCR or whole YAC labeling methods and allows sensitive detection of chimerism.

The large insert size of yeast artificial chromosomes (YACs) has made them the method of choice for cloning and mapping large genomic regions. With the increasing dependence on YACs, however, a number of problems associated with their use have come to light and must be addressed. Principal among these is the high level of chimerism present in several widely used YAC libraries.[1–3] Fluorescence in situ hybridization (FISH) is a convenient and powerful complementary technique to YAC cloning and has the potential to detect chimerism. A number of methods have been described for direct or indirect labeling of YAC recombinants for use in FISH analysis, but current methods are tedious, insensitive, and/or selective with respect to the fraction of the insert labeled.

Labeling total yeast DNA is simple but requires large amounts of DNA and is insensitive. Diverse PCR-based approaches, such as *Alu*–PCR[4] and degenerate oligonucleotide-primed PCR (DOP-PCR),[5] have been developed to facilitate the isolation of discrete internal YAC fragments to generate probes, but all involve tedious efforts to eliminate uneven amplification of the YAC insert and none is comprehensive. Difficulties in obtaining probes representing the entire insert of the YAC also limit the identification of potentially important information within the large clone and restrict the ability to detect chimerism.

Here, we present a simple and efficient method to generate hapten-labeled DNA probes that represent essentially the entire YAC insert using a trace amount of YAC DNA isolated by pulsed field gel electrophoresis (PFGE). This technique generates DNA probes that can be directly used for FISH with high sensitivity and very low background.

MATERIALS AND METHODS

YAC DNA Preparation

YAC DNA was isolated in a gel slice by preparative PFGE in 1% low melting temperature agarose using minimal ethidium bromide and UV light exposure.[6] An equal volume of 0.1 M NaCl was added, and the sample was melted by heating to 70°C for 15 min, then allowed to cool at room temperature for 30 sec. During this 70°C incubation, the sample was mixed vigorously and centrifuged briefly three or four times. One unit of β-agarase I (New England Biolabs) per 50 μl sample volume was added and the reaction incubated at 37°C for 3 hr. YAC DNA was precipitated by the addition of one-tenth volume of 2 M sodium acetate (pH 5.5) and two volumes of ethanol. After chilling at −20°C for 1 hr, the DNA was recovered by 15 min of centrifugation. The DNA pellet was resuspended in 5 μl of *Sau*3AI buffer and digested with 1 unit of *Sau*3AI enzyme at 37°C for 1 hr. The enzyme was inactivated by heating at 70°C for 15 min.

Catch-linker preparation

Oligonucleotides employed were A (5′-CCGAATTCTAGAGTCGACC-3′) and B (5′-GATCGGTCGACTCTAGAATTCGG-

4:209–211©1995 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/95 $5.00

3′). Four micrograms of oligonucleotide B was 5′-phosphorylated by treatment with 25 units of T4 polynucleotide kinase (BRL) in 50 μl at 37°C for 30 min. The oligonucleotide B was recovered by $CHCl_3$/isoamyl alcohol extraction, ether extraction, and ethanol precipitation. Based on an estimated recovery efficiency of 25%, 1 μg of oligonucleotide A was mixed with 1 μg of the kinase-treated oligonucleotide B in 10 μl of 10 mM Tris-HCl, 5 mM $MgCl_2$ (pH 7.4). The mixture was placed in a 200-ml water bath at 70°C and allowed to cool to room temperature over ~1 hr. This produced 200 ng/μl of AB duplexed catch-linker (CL) that can be frozen and stored indefinitely.

CL to YAC Ligation

On ice, ~20 ng of the processed YAC DNA (as judged by the initial pulsed field gel) was mixed with 100 ng of duplexed CL in a 10-μl ligation reaction employing 2 units of T4 DNA ligase (Boehringer Mannheim). After incubation at 16°C overnight, the ligase was heat inactivated at 70°C for 15 min.

PCR amplification and labeling

Oligonucleotide bio-A was synthesized with a single biotin moiety incorporated at the 5′ end of oligonucleotide A during synthesis. Using 600 ng of oligonucleotide A-bio as a primer, 1-μl aliquots of the ligation reaction were amplified in PCR reactions of 50 μl volume using 1 unit of *Taq* polymerase (Boehringer Mannheim) and an OmniGene thermal cycler (Hybaid). PCR conditions were as follows: For 35 cycles using tube temperature control , the denaturation steps were 93°C for 15 sec; the annealing steps were 54°C for 30 sec; and the extension steps were 72°C for 1 min (cycles 1–10), 2 min (cycles 11–20), and 4 min (cycles 21–35). Ten microliters of the PCR product was examined by electrophoresis on a 1.5% agarose gel. The remainder of the PCR reaction was ethanol precipitated and resuspended at a concentration of 20 ng/μl in TE.

Alternative preparation of FISH probes

Yeast DNA containing a human YAC recombinant was isolated by a standard method.[7] *Alu*–PCR was performed according to Breen et al.[4] Both total yeast DNA and *Alu*–PCR products were biotin-labeled by nick translation.

FISH

Chromosomal in situ suppression hybridization was performed as described previously.[8] Briefly, the chromosomal DNA was denatured by incubation of the slides in 0.1 M glycine buffer (pH 12.5), at room temperature for 3 min followed by dehydration in 70% ethanol and 99.5% ethanol at −20°C and then air-dried. The hybridization mixture containing biotin-labeled DNA probes (6 ng/μl for CL+PCR or *Alu*–PCR products, and 100 μg/μl for total yeast DNA), 500 ng/μl of Cot-1 DNA (BRL), 1.5× SSC, 10% dextran sulfate, and 50% formamide were denatured by heating at 78°C for 5 min, chilled on ice, and placed at 37°C for 15 min to allow annealing of repeated DNA sequences and thus suppress their hybridization to chromosomal homologs. Thirty microliters of the hybridization mixture was added to each prewarmed (37°C) slide under a 22×22-mm coverslip, which was sealed with rubber cement. The hybridization was carried out at 37°C overnight in a humid chamber.

Following hybridization, the slides were washed three times in 50% formamide/2× SSC, once in 2× SSC, and once in PN buffer [100 mM sodium phosphate buffer at pH 8.5, with 0.1% NP-40 (Pharmacia)] at 45°C for 5 min each, and then incubated in 5% nonfat dry milk in PN buffer (PNM) at 37°C for 10 min. Signal detection was performed by incubation in FITC–avidin DCS (Vector Laboratories) diluted to 1:100 in PNM at room temperature for 20 min. The signal was amplified with 5 μg/ml of biotinylated anti-avidin antibody D (Vector Laboratories) followed by a second layer of 1:100 FITC–avidin DCS (Vector Laboratories) at room temperature for 20 min each. The successions of antibody incubations were separated by washing twice with PN buffer at room temperature for 2 min each. After the final wash, the slides were mounted by a fluorescent anti-fade solution, VectaShield (Vector Laboratories), containing 0.5 μg/ml of propidium iodide. Chromosome preparations were examined under a Leitz microscope equipped with a Bio-Rad MRC-600 confocal laser scanning system.

Results and Discussion

ICI yac 36IF5, a human YAC recombinant with an insert size of ~600 kb, was identified by hybridization of gridded filters with a microdissection clone, MD220, which we had mapped previously to the long arm of chromosome 11.[9] Figure 1 shows the comparisons of FISH results using probes for ICI yac 36IF5 labeled by three different meth-

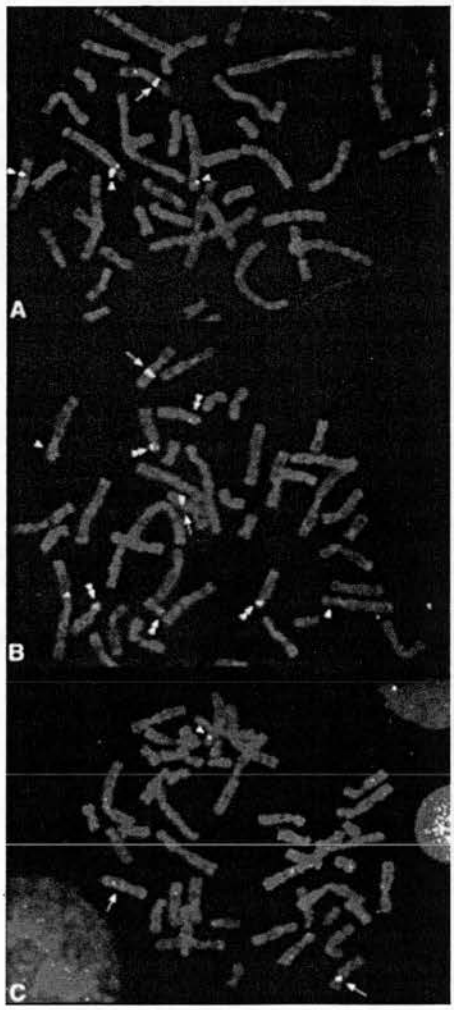


FIGURE 1 Partial metaphases hybridized with ICI yac 36IF5. Probe DNAs are prepared by three different methods: (*A*) (CL)+PCR labeling of ICI yac 36IF5; (*B*) total yeast DNA of ICI yac 36IF5 labeled by nick translation; and (*C*) pooled *Alu*–PCR products for ICI yac 36IF5 followed by nick translation. Arrows indicate FITC signals of ICI yac 36IF5 on chromosome band 11q14.3-q21. This YAC is chimeric and gives minor signals on chromosome band 5p14 (arrowheads). Cross-hybridization of yeast rRNA genes onto acrocentric chromosomes was also seen by the total yeast method (double arrowheads).

ods: (A) CL plus PCR (CL+PCR); (B) total yeast DNA containing a YAC labeled by nick translation; and (C) *Alu*–PCR followed by nick translation.

Arrows indicate FITC signals of ICI yac 36IF5 on chromosome band 11q14.3-q21. This YAC is chimeric as demonstrated by the minor signal on chromosome band 5p14 (arrowheads). The FISH results from both CL+PCR, and the total yeast methods showed clean signals without background noise (Fig. 1A,B). However, YAC DNA consists of only about one-tenth to one-fiftieth of total yeast DNA; therefore, the total yeast method requires a large amount of yeast DNA to compensate for the amount of YAC DNA necessary for FISH. In addition, the vast majority of labeled DNA is the yeast genomic DNA, and this causes cross-hybridization of conserved sequences such as rRNA onto human chromosomes (Fig. 1B: double arrowheads).

All three labeling methods correctly localize ICI yac 36IF5 to 11q14.3-q21. The study, however, illustrated several points that limit the usefulness of total yeast or *Alu*–PCR and favor the CL+PCR method.

1. *Alu*–PCR labeling is accompanied by a high background noise that may obscure or falsely indicate chimerism (Fig. 1C).
2. Although both CL+PCR and total yeast labeling give low background noise, the amount of material required for total yeast labeling is large and the sensitivity is low. As little as 2 ng of YAC will generate by CL+PCR sufficient probe for 10 FISH analyses.
3. Labeling of total yeast DNA but not CL+PCR results in labeling of human homologs of yeast sequence, most notably rRNA genes (Fig. 1B, double arrowheads).
4. The success of *Alu*–PCR (or related methods) but not CL+PCR is limited to human recombinants and is heavily dependent on the local distribution of repeated DNA sequence motifs that serve to prime PCR.
5. Because all ends of restriction fragments are converted to priming sites for PCR in CL+PCR, but not other methods, essentially the complete YAC DNA is directly available without sequence bias for FISH, comprehensive cloning, and further molecular analysis by other techniques.

Because the DNA fragment size range after CL+PCR technique (200–700 bp; data not shown) is optimal for FISH without further digestion with DNase I, hapten labeling of DNA fragments during the PCR reaction is preferred. The conventional way of labeling DNA during PCR is to substitute 1 of the 4 nucleotides in the reaction mixture with a labeled one such as bio-*n*-dUTP or DIG-11-dUTP. This requires a high concentration of labeled nucleotide, and is therefore costly. On the other hand, the labeling efficiency by CL+PCR using a biotinylated primer without additional labeled nucleotide in the PCR reaction is adequate both for the standard signal detection procedure for FISH (Fig. 1A) and for filter hybridization (data not shown). Our experiences with CL+PCR have shown the method to be extremely reliable and sensitive (12 YACs tested and all successful). Estimated from the results of contig construction, the size of chimeric fragment of ICI yac 36IF5 mapped on chromosome 5 is <50 kb. Furthermore, once the YAC clone is confirmed to be valuable for further analysis, the amplified DNA fragments by CL+PCR can be subcloned into suitable vector, using the internal *Eco*RI site on the linker adapter or the *Sau*3AI of the original ligation site.

We conclude that the CL+PCR labeling technique offers valuable advantages over current methods for high-resolution YAC mapping and the detection of chimerism by FISH analysis.

ACKNOWLEDGMENTS

We thank the Photography Department, MRC Human Genetics Unit, Edinburgh, for preparing Figure 1. This study was supported in part by grants from the U.K. Human Genome Mapping Project (grants E321-118 and E321-159).

REFERENCES

1. Selleri, L., J.H. Eubanks, M. Giovannini, G.G. Hermanson, A. Romo, M. Djabali, S. Maurer, D.L. McElligott, M.W. Smith, and G.A. Evans. 1992. Detection and characterization of "chimeric" yeast artificial chromosome clones by fluorescent *in situ* suppression hybridization. *Genomics* **14:** 536–541.
2. Bates, G.P., J. Valdes, H. Hummerich, S. Baxendale, D.L. Le Paslier, A.P. Monaco, D. Tagle, M.E. MacDonald, M. Altherr, M. Ross, B.H. Brownstein, D. Bentley, J.J. Wasmuth, J.F. Gusella, D. Cohen, F. Collins, and H. Lehrach. 1992. Characterization of a yeast artificial chromosome contig spanning the Huntington's disease gene candidate region. *Nature Genet.* **1:** 180–187.
3. Shibasaki, Y., J.C. Maule, R.S. Devon, E.M. Slorach, J.R. Gosden, D.J. Porteous, and A.J. Brookes. (unpubl.).
4. Breen, M., B. Arveiler, I. Murray, J.R. Gosden and D.J. Porteous. 1992. YAC mapping by FISH using *Alu*–PCR-generated probes. *Genomics* **13:** 726–730.
5. Telenius, H., N.P. Carter, C.E. Bebb, M. Nordenskjöld, B.A.J. Ponder, and A. Tunnacliffe. 1992. Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. *Genomics* **13:** 718–725.
6. Maule, J.C., D.J. Porteous, and A.J. Brookes. 1994. An improved method for recovering intact pulsed field gel purified DNA, of at least 1.6 megabases. *Nucleic Acid Res.* **22:** 3245–3246.
7. Hoffman, C.S. and F. Winston. 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* **57:** 267–272.
8. Shibasaki, Y. 1994. High resolution mapping of the MYCN proto-oncogene at human chromosome 2p24.3 by fluorescence *in situ* hybridisation. *Cytogenet. Cell Genet.* **66:** 75–76.
9. Muir, W.J., C.M. Gosden, A.J. Brookes, J. Fantes, K.L. Evans, S.M. Maguire, B.J. Stevenson, S. Boyle, D.H.R. Blackwood, D. St. Clair, D.J. Porteous, and A. Weith. 1995. Direct microdissection and cloning of a translocation breakpoint region t(1;11)(q42.2;q21) associated with schizophrenia. *Cytogenet. Cell Genet.* (in press).

Received August 29, 1994; accepted in revised form December 2, 1994.

Molecular Biotechnology (in press)

Coincidence Cloning: taking the coincidences out of genome analysis

Rebecca S. Devon and Anthony J. Brookes

Abstract

The term Coincidence Cloning encompasses a wide range of methodologies, the aim of which is to isolate DNA sequences which occur in both of two input DNA sources. The nature of these input DNAs may be genomic or cDNA, cloned or uncloned, and as such the far reaching applicability of the techniques can be imagined. If the input DNAs are genomic then the product will be enriched for useful markers co-occurring between the two. If the input DNAs comprise one genomic resource and one cDNA resource the product will contain genes mapping to that particular genomic region. In this review a comparative description of the range of Coincidence Cloning methods is given, together with a discussion of their applications. Finally, consideration is given to the general limitations of these techniques.

Key Words

cloning, coincident, methods, PCR, enrichment

Despite the many and various molecular genetic technologies currently available, a detailed analysis of all but the smallest genomes remains a colossal and extremely arduous task. The primary reason for this is the sheer size and complexity of the genomes currently being scrutinised. To help counter this problem, efforts have been made to develop strategies for DNA investigation that are ever more sophisticated and increasingly powerful. Several such procedures share, as a common purpose, the selective cloning of DNA sequences that are distinguished by their presence in both of a pair of DNA resources. By and large, these methods are dependent upon the polymerase chain reaction (PCR) (1). A generic name that may be applied to these technologies is 'Coincidence Cloning' (the cloning of sequences *coinciding* between two resources).

Systems for Coincidence Cloning have been developed for use with many types of DNA resource. These include i) genomic DNA mixtures such as individual or pooled recombinants (cosmids, phage, yeast artificial chromosomes etc.) and even whole somatic cell hybrid or single species total genomic DNAs, ii) the transcribed portions of a genome available in the form of cDNA resources, and iii) pre-constructed libraries of whole genomes, sub-genomes or tissue specific cDNAs. The application of Coincidence Cloning to a pair of genomic DNAs will provide access to a highly focused set of genomic markers. Alternatively, if one employs a genomic resource plus a cDNA resource then the procedure will identify groups of transcribed sequences encoded within that genomic region. These fundamental goals of marker or gene isolation represent the current applications of Coincidence Cloning and hence are the focus of this review. However, other possibilities for exploitation of the Coincidence Cloning concept can be imagined, and are considered briefly.

For convenience we shall refer to the DNA mixtures used in coincidence cloning studies as Source I and Source II DNAs, use 'integration' to describe their joint processing and the 'product' to describe the resultant set of coincident DNAs. In most circumstances, Coincidence Cloning procedures do not actually purify target sequences to homogeneity, instead coincident DNAs are preferentially recovered. The degree of enrichment achieved is expressed in terms of an 'enrichment factor'. The enrichment factor is given by the ratio of the relative mass representation of a target sequence in the product DNA mixture, over its relative mass representation in one of the (usually the most complex) input DNA resources.

Coincidence Cloning procedures are all based upon the same underlying principle, namely the formation and subsequent selective isolation of duplex DNA molecules (Inter-Resource Duplexes; IRDs) comprised of complementary DNA strands from each of the input DNA resources. This general scheme is represented in figure 1. The logic of this scheme is clear, since IRDs can exist only for coincident DNA sequences. Thus, Coincidence Cloning can be viewed simply as a form of DNA hybridisation that uses a highly complex probe to both detect and simultaneously recover the hybridising sequences of interest. The fundamental challenge to coincidence

cloning lies not in the hybridisation step (the kinetics of duplex DNA formation is well understood) but in the selective recovery of the desired sequences (the IRDs).

For the purposes of this review, the primary modes that have been used for the IRD selection have been divided into four groups. These are i) cloning selection, ii) physical selection, iii) physical plus PCR selection, and iv) physical plus PCR plus ligation selection. A summary of these selection methods is shown in Table 1. The basis for Cloning Selection is the selective cloning of the *inter*-resource duplexes but not the *intra*-resource duplexes. This is usually achieved by the presence of different restriction enzyme sites on each end of the IRD molecule. Although the use of cloning as a selection method removes the need for subsequent cloning of the product prior to analysis, non-specific products are usually recovered at an unacceptably high level. Physical Selection would be the method of choice for Coincidence Cloning studies employing whole chromosomes. The Source II DNA is used as a fluorescent *in-situ* hybridisation (FISH) probe, which enables visualisation of the region of coincidence before physically removing the bound probe from the slide. This is useful for large scale studies, although the requirement for whole chromosomes on a slide reduces its versatility. Physical plus PCR Selection makes use of different primer sites incorporated into the Source I and II DNAs, along with physical retention of one of the input DNAs. This allows molecules not forming a stable duplex to be washed away, and then the IRD may be selectively amplified by PCR. PCR selection can achieve greater enrichment than either cloning or physical selection alone. However, the product may contain non-coincident products which had not been removed by the washing steps, and PCR itself is subject to a number of limitations (as discussed later). An improvement on straightforward PCR selection, Physical plus Ligation plus PCR Selection, involves ligation of primer sites onto the Source II product molecules while they are present in a duplex with Source I DNA. Stringent conditions in the ligation step ensure that only truly coincident molecules will acquire the primer sites essential for amplification and selection. The enrichment factors achieved by the incorporation of this extra step are higher than with standard PCR selection.

Two major factors conspire to compromise the effectiveness of coincidence cloning. The first of these is input DNA complexity. A typical mammalian genome is so large (~10^6 kb) that to integrate anything but relatively small portions of it would require unacceptably long incubation times to achieve effective DNA hybridisation. To overcome this, most procedures simplify the Source I and/or Source II DNAs, usually via a PCR step, to create 'representations' of the originals. The second problem is that all repetitive elements present in both source DNAs will be genuinely coincident and therefore recovered in the product DNA. For example the human Alu repeat family is highly reiterated and members comprise approximately 5% of the total genome and are present in 3-5% of cDNAs (2). This high frequency means that there will be Alu repeats coincident between the two resources which therefore will be enriched in the product library. However, pre-

hybridisation of the denatured input DNAs with an excess of Alu repeat DNA before duplex formation can result in the near complete functional removal of the Alu repeat sequences from the input DNAs (3). Similarly, this 'pre-blocking' can also be used to alleviate other problems of non-specific recovery, such as for ribosomal RNA sequences (3).

The methodologies which fall under the general heading of Coincidence Cloning can be distinguished by the different ways in which the IRD is selected. In order to differentiate the *inter*-resource duplex from the *intra*-resource duplex, it is necessary to distinguish between the Source I and the Source II DNAs in some manner. This may involve modification of the ends of the molecules by addition of different linkers to each resource, such that they contain different restriction enzyme sites or PCR primer sites, or alternatively differentiation may simply involve physical separation of one from the other by binding to a filter or fixing on a slide. Once the two Source DNAs can be discriminated in this manner, then one of the selection methods can be employed.

Methods for selecting the IRD

Cloning selection

For the IRDs to be selected by cloning, use can be made of different restriction enzyme sites in the linkers used to amplify the starting DNA (4,5). This is depicted in figure 2. The Source I is amplified by PCR using a primer that incorporates a restriction enzyme site. The Source II is similarly amplified, but a different enzyme site is incorporated into the otherwise similar primer. The two resulting PCR products are then digested with the appropriate restriction enzyme, before denaturing and mixing them together. As reannealing takes place, the Source I and Source II sequences may of course reanneal to themselves, yet IRDs will also be formed and are distinguishable by the presence of different restriction enzyme half-sites at each end of the duplex. These may then be directly selected by cloning all the duplexed material into an appropriately doubly digested vector. Only the IRD formed between source I and source II DNAs would be able to be correctly ligated into the vector and cloned.

Cloning selection was utilised in an early coincidence cloning experiment, in which markers were isolated from two somatic cell hybrids which contained a small amount of human DNA in common (4). Alu-PCR (6) was used to amplify the somatic cell hybrids, employing Alu consensus primers for each hybrid incorporating a different restriction enzyme site. This study achieved an enrichment factor of 10-20 fold. The use of repeat directed PCR with cloning selection has also been extended to other species. For instance primers directed towards a pig specific SINE have been used to amplify the porcine component of a somatic cell hybrid (the Source I DNA) and a flow sorted pig chromosome 6 (the Source II DNA), with subsequent cloning selection employed to enrich for the coincident sequences between them (7).

Cloning selection is limited by the recovery of significant levels of background products due to i) the cloning of Source I-only or Source II-only duplexes into an incompletely digested vector, ii) the cloning of chimaeric products and iii) the recovery of false IRDs that have poor inter-strand homology. For these reasons other methods for selection of the IRDs have since been developed.

Physical selection

The techniques of FISH have been applied to coincidence cloning, allowing visualisation of the region of coincidence before physical isolation of the IRD. This approach is useful when the experiment is on a large scale, for instance when one or more of the input DNAs comprises whole chromosomes. Figure 3 depicts a general scheme for physical selection. The Source I DNA is used as a FISH probe onto the Source II DNA which has been fixed on a slide. Excess probe and that which is non-specifically bound can be washed off, enabling the genuinely coincident sequences to be visualised and then eluted off, PCR amplified and cloned.

The two methods of 'Coincidence painting' (8) and 'Prep-ISH' (9) employ FISH and physical selection methods. Coincidence Painting was used to obtain marker sequences from the region of coincidence between two chromosomes which share a small portion of their sequence - the two DNA sources being a derivative chromosome from a translocation cell line plus flow sorted normal chromosomes. It was calculated that this method yielded approximately one unique clone for every 53kb of the coincident sequence. This is a substantially higher density of markers than that produced by the Alu-PCR based method, which produced on average one clone every 250kb (4).

In contrast, Prep-ISH (9) is a method published primarily to isolate genes mapping to a specific chromosomal region although it could equally be applied to isolating coincident genomic markers. A linker-amplified cDNA library or genomic region of interest is used as a probe for FISH onto an array of chromosomes fixed on a slide. After appropriate washing, the region of coincidence is recovered by microdissection and reamplification. Mammalian genomes can be resolved into 300 to 1000 or more recognisable chromosome bands, thus a cDNA library containing 10,000 different transcripts could be reduced in complexity by Prep-ISH to a more manageable 10 to 30 transcripts mapped to a specific region of the genome. This is potentially extremely useful for producing a small set of candidate genes from a specific genomic region without the need for further physical mapping and selection of cDNAs.

Physical plus PCR selection

In solely physical selection methods, the elution of the bound probe from the slide provides separation of the IRD from the background, and yet for there to be sufficient product for analysis it is necessary to first PCR amplify the eluted material. An expedient alternative therefore would be a PCR based

selection approach, which could select the IRD and concomitantly result in sufficient material for analysis.

In methods employing PCR selection (see figure 4), oligonucleotide linkers of different sequence are used to amplify the Source II and usually the Source I material. In addition, the denatured Source I DNA is attached to a physical support, which may be either by binding to a filter or by attachment to streptavidin coated magnetic beads. After integration of the two resources, the IRD is physically retained by attachment to the filter or beads, while the excess Source II DNA is removed by washing. The product Source II can then be eluted off the physical support and amplified using specific primers. This strategy is therefore closely related to the FISH based methods described above, but is more versatile.

PCR selection has been largely used for gene isolation studies, whereby the product is enriched in cDNAs mapping to a particular genomic region (i.e. coincident between a cDNA resource and a specific genomic resource). This technique is sometimes referred to as 'cDNA fishing', with the analogy that the genomic region acts like a fishing rod, to dip into the pool of cDNA 'fish' and physically reel in coincident sequences.

Filter hybridisation The first experiments using PCR selection were related schemes for cDNA fishing, called 'direct cDNA selection'. These involved the hybridisation of a whole cDNA library (the Source II) onto YACs or cosmids (the Source I) which had been immobilised on a filter. In one study (10) an amplifiable foetal kidney cDNA library was pre-blocked for high copy repeats and integrated with a 550kb YAC, which was denatured and immobilised on a filter. After washing off the non-specifically bound cDNAs the product was eluted off, amplified and cloned. In the second study (11) the only fundamental difference was that the immobilised genomic DNA, not the cDNA, was pre-blocked for repeats. The optimisation of hybridisation and washing conditions is obviously critical in this type of experiment, to determine the level of specificity achieved. The enrichment factor achieved in both studies was ~1000 fold. This is impressive yet is still insufficient to reliably recover rare transcripts.

Solution Hybridisation In later studies the focus of Coincidence Cloning experiments has changed to solution rather than filter hybridisation (12,13,14), for which the kinetics are more easily controlled. As an alternative to immobilisation of the genomic DNA on a filter, the genomic DNA is usually attached to streptavidin-coated magnetic beads by way of biotin moieties. The biotin residue may be added during the initial amplification of the genomic DNA resource, either by the use of a biotinylated primer or one biotinylated nucleotide in the reaction mix. The pre-blocked amplified cDNA (Source II) is added before (14) or after (12) the genomic DNA fragments (Source I) have been immobilised. cDNAs forming stable duplexes with the genomic DNA will be retained by the beads upon washing, while the excess cDNA is removed. The product cDNA may then be eluted, amplified and cloned as before. Further rounds of cDNA selection can be applied to

increase the enrichment, although at the possible expense of some genuinely coincident cDNA species, since further rounds of PCR amplification can induce biases in sequence recovery.

Physical plus Ligation plus PCR selection

In PCR selection methods, it is straightforward to separate Source II product DNA from Source I DNA, since the primers used to amplify each resource are not compatible with each other. However, the primers used to amplify the product Source II DNA are equally able to amplify Source II DNA from imperfectly matched IRDs and also Source II *intra*-resource duplexes which have not been removed during the washing steps. This can lead to the recovery of artefacts in the product material.

A method termed 'end ligation coincident sequence cloning' (EL-CSC)(15,16) has been developed which, by incorporating an extra step into the protocol significantly reduces the contamination of the product library with non-coincident Source II DNA. This is shown in figure 5. The extra step is a high stringency ligation reaction which joins synthetic 'capture' oligonucleotides onto the Source II DNA molecules while they are present in a duplex with Source I molecules. The capture oligonucleotides are designed to be complementary to linkers present at the ends of the Source I molecules and hence the ligation can only take place if the Source I and Source II IRD is perfectly matched both in length and terminal sequence. After washing and elution, the product Source II molecules are amplified using primers directed against the unique regions of the added oligonucleotides. Hence only those coincident products which fulfil the criteria of perfect length match and end sequence identity will be amplified and subsequently cloned. The technique has been validated in studies designed either to isolate region specific markers (single copy markers coincident between somatic cell hybrids and a pool of microdissection clones) (15) or transcribed sequences from a particular genomic region (six genes were isolated as coincident between 260kb of pooled cosmids and human foetal brain cDNA) (15).

The higher specificity of EL-CSC relative to cDNA fishing techniques means that enrichment factors of greater than 10^6 fold can be achieved, at least 10 fold greater than those possible with most cDNA fishing. It has been found however that a slightly different spectrum of products is generated by performing EL-CSC versus cDNA fishing with the same input DNAs. The most efficient way of detecting coincident sequences would thus appear to be the use of a combination of EL-CSC and cDNA fishing, with the two methods being viewed as complements rather than alternatives. With this objective in mind, a procedure has now been developed which can accomplish these two objectives in a single tube reaction (17).

General problems and limitations

Artefactual products

Inevitably, some of the DNA present in the final product will be artefactual sequences. Low copy genomic repeat elements such as MER sequences can confound the search for real genes, since they are sometimes present in expressed elements, making them genuinely coincident between the two resources (3). Their copy number is insufficiently high to permit effective blocking in the preannealing step even by Cot10 DNA. Also, the design of cDNA fishing means that in addition to the genuine gene, other members of a gene family or expressed pseudogenes may well have sufficient cross-homology to be fished out.

These problems have been overcome to a certain extent by EL-CSC since the requirements for perfect length match and end-sequence identity will less frequently be met in pseudogenes and related sequences. In this respect, EL-CSC may be seen to a greater extent as a method for *isolating* genes rather than just enriching them. A constraint of EL-CSC is that it is only able to isolate intra-exonic fragments due to the requirement for length match between the genomic and cDNA versions of the gene. However, inspection of a random set of gene sequences for which the genomic structure is known reveals that the majority will contain multiple recoverable fragments using one or more of a set of common restriction enzymes (15).

PCR

PCR amplification is an essential component of most coincidence cloning techniques, both those involved in marker isolation and in gene isolation. PCR amplification is generally used in the preparation of the input DNAs (for instance Alu-PCR (6), DOP-PCR (18) and catch linker PCR (19) and is also necessary for the amplification of the coincident product sequences, so that there is sufficient material for cloning and further manipulation. It is therefore unfortunate that PCR amplification does not result in an unadulterated expansion of the starting material. Firstly, PCR amplification is biased towards shorter fragments, so it is important to ensure that the starting material is digested to a suitable length. Standardising the length of material to be amplified also helps to ensure that it is all amplified - longer fragments would be in direct competition with the shorter ones and invariably lose. Secondly, PCR biases are also prevalent against certain sequences with high G+C content, resulting in uneven amplification of the starting material. Some length and sequence biases may be overcome by modification of the PCR programme, for instance by employing longer extension times and sufficiently high temperatures such that GC-rich sequences are properly denatured. A third drawback associated with PCR is that sequence errors are introduced by the polymerase, which may corrupt open reading frames and give other artefactual results. High fidelity polymerases with a proof-reading activity have been recently introduced, which should reduce this problem. Minimising the number of PCR cycles is important in reducing all three problems.

Quality of input DNA resources

The effectiveness of Coincidence Cloning at enriching a product library for sequences of interest is critically dependent upon the methodology used. Equally important however is the type and quality of the input DNAs. The ability of coincidence cloning to handle complex genomic and cDNA resources obviously results in a limitation in efficiency. In this respect, fewer artefactual products are usually recovered when using cosmid DNA rather than YAC DNA (3), cosmids being approximately one tenth of the size of YACs and more readily produced in a pure form. A balance must be achieved between the level of specificity required for a particular experiment against the length of the genomic region to be studied.

a) YACS A significant drawback when YACs are used as the genomic Source I resource is that they are difficult to isolate in a pure form - excised preparative pulse field gel slices will inevitably be contaminated with yeast sequences to some degree (3). Parts of the yeast genome show sufficient conservation with human cDNA that they may be genuinely recovered by coincidence cloning. Ribosomal RNA sequences are particularly noteworthy in this respect due to their high abundance in the yeast genome. Pre-blocking the cDNA with an excess of denatured ribosomal RNA however can significantly alleviate the problem. In the same way, mitochondrial DNA and 2μ plasmid sequences in the cDNA can be problematic when using cDNA libraries that may contain these elements (3).

Due to the difficulty in isolating the pure YAC DNA, some experimenters have used total yeast DNA plus YAC as the genomic resource. Parimoo et al (1993) isolated HLA-A cDNA clones in this way (20), although the practical advantage did result in a drop in sensitivity. Enrichment of only 100 fold was achieved after two rounds of selection. Forster and Rabbitts (1993) also used the total yeast plus YAC method to isolate fusion cDNAs from tumour tissue (21), and achieved a 10^4 fold enrichment after two rounds of selection. Amplification of a total yeast plus YAC resource by Alu-PCR would selectively amplify the human YAC sequences rather than yeast sequences, although as discussed earlier, PCR biases and non-random distribution of Alu sequences would result in a genomic resource not truly representative of the starting material.

b) cDNA The quality of the cDNA resource in gene isolation experiments is also of crucial importance. It is generally found that oligo-dT primed cDNA gives the best results, since cDNA libraries are often plagued with artefacts, and problems can arise due to homology between the cDNA and genomic resource vectors resulting in vector sequences forming a large proportion of the product material. A potential problem with oligo-dT primed cDNA can be foreseen if the region of coincidence lies solely at the 5' end of a long cDNA, since amplification of this 5' end will be at a disadvantage in a PCR reaction. However the observed greater exonic length at the 3' end is advantageous, especially for EL-CSC. Alternatively, the cDNA may have a long internal hairpin structure, which disables it from forming a duplex with coincident genomic DNA. The tissue from which the cDNA is made must be carefully

selected, since genes not expressed in this tissue cannot be recovered by Coincidence Cloning no matter how high the enrichment factors. To partially overcome this doubt, it may be preferable to use a cDNA resource which has been made from a mixture of tissues, each with its overlapping but slightly different complement of expressed genes. This insurance policy may however result in the non-recovery of genes rarely expressed in only one tissue, since the recovery rate of genes will be dependent on their expression level. If developmental genes are involved, non-recovery may be simply due to the fact that the genes are only expressed (and hence only present in the cDNA resource) at particular developmental stages. This can be somewhat problematic to overcome for human genes, however if there is sufficient cross species conservation use could potentially be made of mouse developmental stage-specific cDNA.

Future Directions

But what of the future of Coincidence Cloning? The need for isolation of marker sequences from the human genome is not so pressing as it was several years ago, and the pace of gene isolation is stepping up with the development of improved technologies in fields such as exon trapping (22,23) and the increasingly realistic concept of direct sequencing of large genomic regions coupled with effective gene finding algorithms (24,25). For these reasons, the flexibility of coincidence cloning and its ability to use large uncloned genomic regions as input DNA resources are to its distinct advantage. The ability of coincidence cloning to scan long stretches of DNA for genes means that candidate gene analysis may begin earlier in the process of positional cloning, before fine scale mapping and linkage analysis have maximally narrowed down the region of interest.

One potential avenue for expansion and development of the technology is in the identification of conserved sequences between species. Just as a conventional 'zoo blot' can reveal the genic nature of a small probe by assessing its cross-species conservation, Coincidence Cloning can do the same, but on a much larger scale. Preliminary experiments in this field have been reported (26) in which conserved regions from the human genome were isolated by performing coincidence cloning between human cosmids and mouse or pig genomic clones, by using a low stringency magnetic bead based protocol.

The ultimate extension to finding regions of coincidence between two subgenomic regions or cDNAs is that whereby whole human genomes become the input DNA sources. The concept of 'Identity by Descent' (IBD) may then be exploited, whereby among consanguineous individuals affected with a genetic defect, the only region of complete genomic sequence identity will be that involved in causing the disease, (the location of a disease gene)(27). On a small scale, IBD has been demonstrated by genotyping microsatellite markers in individuals affected by Hirschprung disease (28). It is not difficult to extrapolate this idea and to visualise an application for high

stringency coincidence cloning in a larger scale search where the location of the disease gene has not been pinpointed.

Acknowledgements

We would like to thank Prof. David Porteous for his critical comments on the manuscript.

References

1) Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis,K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487-491.

2) Crampton, J.M., Davies, K.E. and Knapp, T.F. (1981) The occurrence of families of repetitive sequences in a library of cloned cDNA from human lymphocytes. *Nucleic Acids Research*, **9**, 3821-3834.

3) Lovett, M. (1994) Fishing for complements: finding genes by direct selection. *Trends in Genetics*, **10**, 352-357.

4) Aslanidis, C. and de Jong, P.J. (1991) Coincidence Cloning of Alu-PCR products. *Proc. Natl. Acad. Sci. USA.*, **88**, 6765-6769.

5) Brookes, A.J. and Porteous, D.J. (1992) Coincident sequence cloning: a new approach to genome analysis. *TibTech*, **10**, 40-44.

6) Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.D., Ledbetter, D.H. and Caskey, C.T. (1989) *Alu* polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. USA.*, **86**, 6686-6690.

7) Frengen, E., Thomsen, P.D., Schmitz, A., Frelat, G. and Davies, W. (1994) Isolation of region specific probes from pig chromosome 6 by coincidence cloning. *Mammalian Genome*, **5**, 497-502.

8) Bailey, D.M.D., Carter, N.P., de Vos, D., Leversha, M.A., Perryman, M.T. and Ferguson-Smith, M.A. (1993) Coincidence Painting: a rapid method for cloning region specific DNA sequences. *Nucleic Acids Research*, **21**, 5117-5123.

9) Hozier, J., Graham, R., Westfall, T., Siebert, P. and Davis, L. (1994) Preparative *in situ* hybridisation: selection of chromosome region-specific libraries on mitotic chromosomes. *Genomics*, **19**, 441-447.

10)Lovett, M., Kere, J. and Hinton, L.M. (1991) Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. USA.*, **88**, 9628-9632.

11)Parimoo, S., Patanjali, S.R., Shukla, H., Chaplin, D.D. and Weissman, S.M. (1991) cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci. USA.*, **88**, 9623-9627.

12)Morgan, J.G., Dolganov, G.M., Robbins, S.E., Hinton, L.M. and Lovett, M. (1992) The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. *Nucleic Acids Research*, **20**, 5173-5179.

13)Abe, K. (1992) Rapid isolation of desired sequences from lone linker PCR amplified cDNA mixtures: application to identification and recovery of expressed sequences in cloned genomic DNA. *Mammalian Genome*, **2**, 252-259.

14)Korn, B., Sedlacek, S., Manca, A., Kioschis, P., Konecki, D., Lehrach, H. and Poustka, A. (1992) A strategy for the selection of transcribed sequences in the Xq28 region. *Human Molecular Genetics*, **1**, 235-242.

15)Brookes, A.J., Slorach, E.M., Morrison, K.E., Qureshi, S.J., Blake, D., Davies, K. and Porteous, D.J. (1994) Cloning the shared components of complex DNA resources *Human Molecular Genetics*, **3**, 2011-2017.

16)Brookes, A.J. (1994) Identifying and directly purifying transcribed elements by coincident sequence cloning. *In 'Identification of Transcribed Sequences' (ed. Hochgeschwender, U. And Gardiner, K.),* Plenum Press, New York.

17)Morrison, K.E., Qureshi, S.J., Anderson, S., Borrett, J.P., Theodosiou, A., Blake, D., Nesbit, A., Davies, K.E., Porteous, D.J. and Brookes, A.J. (1995) Novel genes from the complex genomic region 5q13, identified by end-ligation coincident sequence cloning. *Genomics*, submitted.

18)Telenius, H., Pelmear, M., Tunnacliffe, A., Carter, N.P., Behmel, A., Ferguson-Smith, M.A., Nordenskjold, M., Pfragner, R. and Ponder, B.A. (1992) Cytogenetic analysis by chromosome painting using DOP-PCR amplified flow sorted chromosomes. *Genes-Chromosomes-Cancer*, **4**, 257-263.

19)Shibasaki, Y., Maule, J.C., Devon, R.S., Slorach, E.M., Gosden, J.R., Porteous, D.J. and Brookes, A.J. (1995) Catch-linker + PCR labelling: a simple method to generate fluorescence in situ hybridisation probes from yeast artificial chromosomes. *PCR Methods and Applications*, **4**, 209-211.

20)Parimoo, S. Kolluri, R. and Weissman, S.M. (1993) cDNA slection from total yeast DNA containing YACs. *Nucleic Acids Research*, **21**, 4422-4423.

21)Forster, A. and Rabbitts, T.H. (1993) A method for identifying genes within yeast artificial chromosomes: application to isolation of *MLL* fusion cDNAs from acute leukaemia translocations. *Oncogene*, **8**, 3157-3160.

22)Dyuk, G.M., Kim, S., Myers, R.M. and Cox, D.R. (1990) Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl.Acad. Sci. USA*, **87**, 8995-8999.

23)Buckler, A.J., Chang, D.D., Graw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A. and Housman, D.E. (1991) Exon amplification: a strategy to isolate mamalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA*, **88**, 4005-4009.

24)Smith, M.W., Holmsen, L.N., Wei, Y.H., Peterson, M. and Evans, G.A. (1994) Genomic sequence sampling: a strategy for high resolution sequence-based physical mapping of complex genomes. *Nature Genetics*, **7**, 40-47.

25)Claverie, J-M. (1994) A streamlined random sequencing strategy for finding coding exons. *Genomics*, **23**, 575-581.

26)Lander, E.S. and Botstein, D. (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*, **236**, 1567-1570.

27)Sedlacek, Z., Konecki, D.S., Siebenhaar, R., Kioschis, P. and Poustka, A. (1993) Direct selection of DNA sequences conserved between species. *Nucleic Acids Research*, **21**, 3419-3425.

28)Puffenberger, E.G., Kauffman, E.R., Bolk, S., Matise, T.C., Washington, S.S., Angrist, M., Weissenbach, J., Garver, K.L., Mascari, M., Ladda, R., Slaugenhaupt, S.A. and Chakravarti, A. (1994) Identity-by-descent and association mapping of a recessive gene for Hirschprung disease on human chromosome 13q22. *Human Molecular Genetics*, **3**, 1217-1225.

Table 1) Methods for the selection of inter resource duplex (IRD) DNA

Selection Method	Principle	Advantage	Disadvantage
1) Cloning	Directed cloning	Removes need for cloning step after selection	High level of background products
2) Physical	FISH hybridisation and physical isolation	Useful for large scale studies	Requires fixed, metaphase chromosomes
3) Physical plus PCR	Physical isolation and selective PCR	High level of enrichment	Potential background products
4) Physical plus Ligation plus PCR	Physical isolation, ligation of primer sites and selective PCR	Very high level of enrichment	Requirement for identical restriction fragments between Source DNAs

Fig. 1) General schema for coincidence cloning

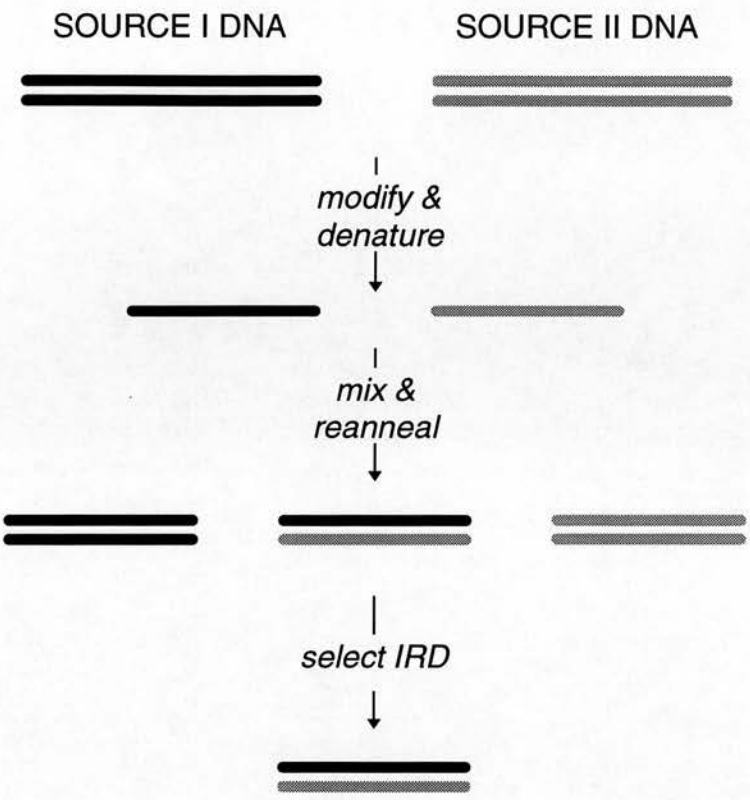


Fig. 2) Cloning selection

■ & □ represent different restriction enzyme sites incorporated into the primers

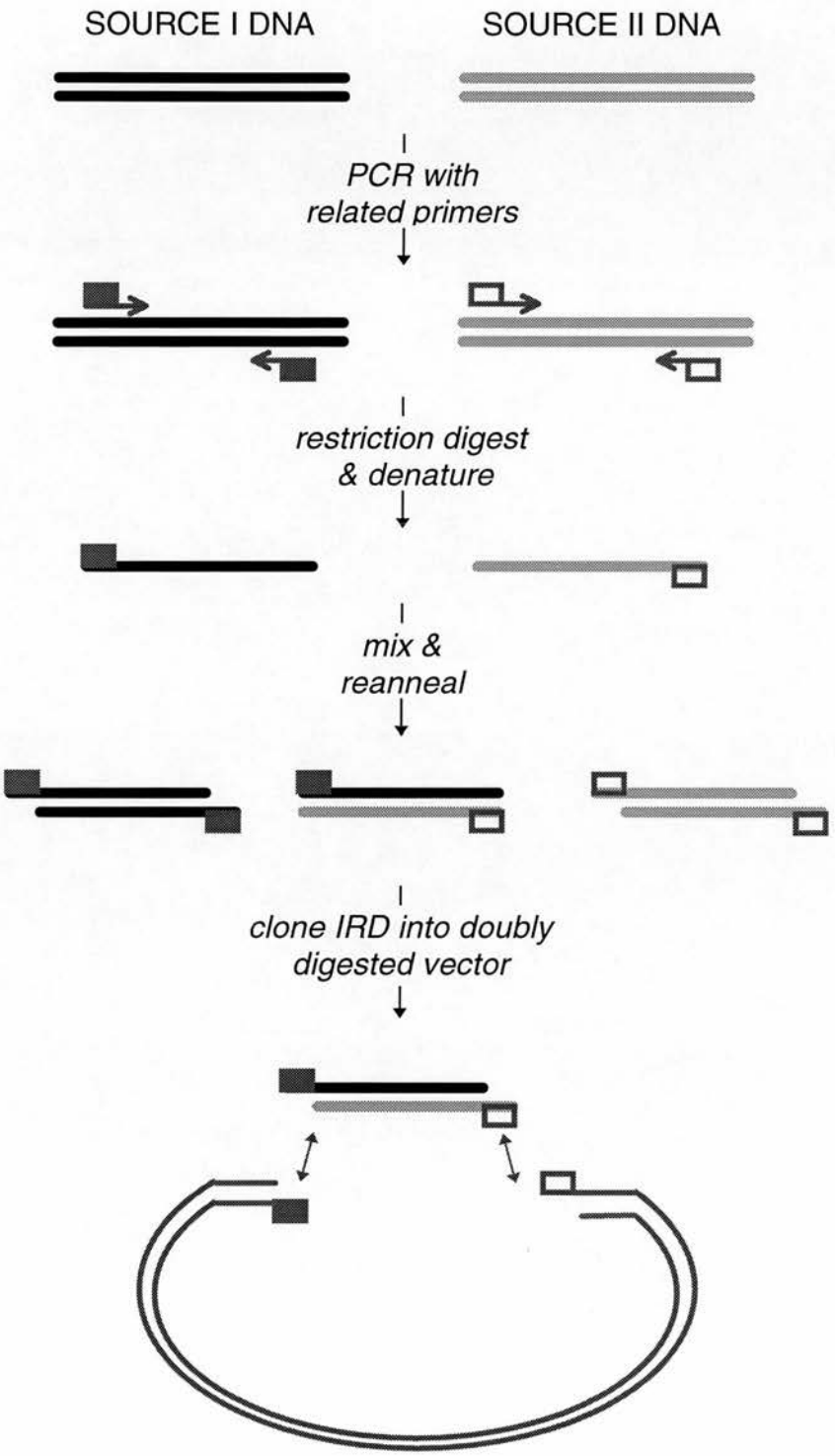


Fig. 3) Physical selection after FISH

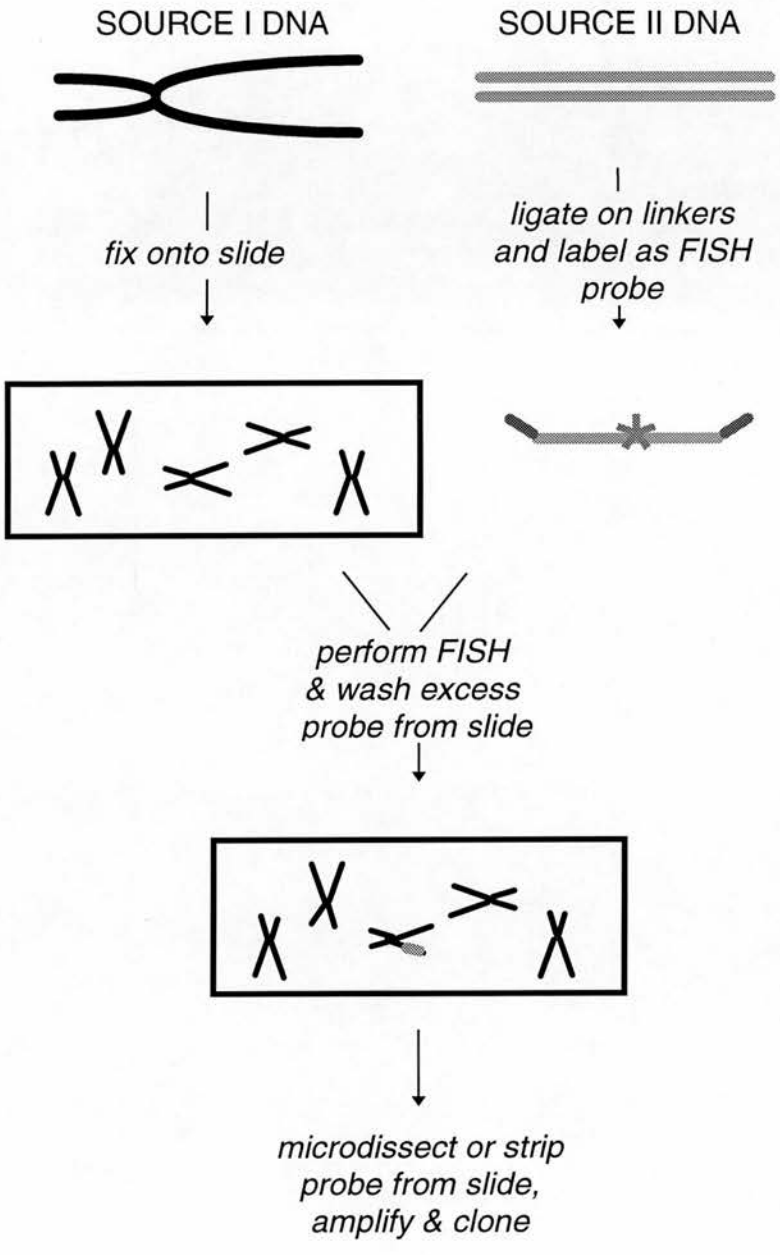


Fig. 4) Physical plus PCR selection

B = biotin

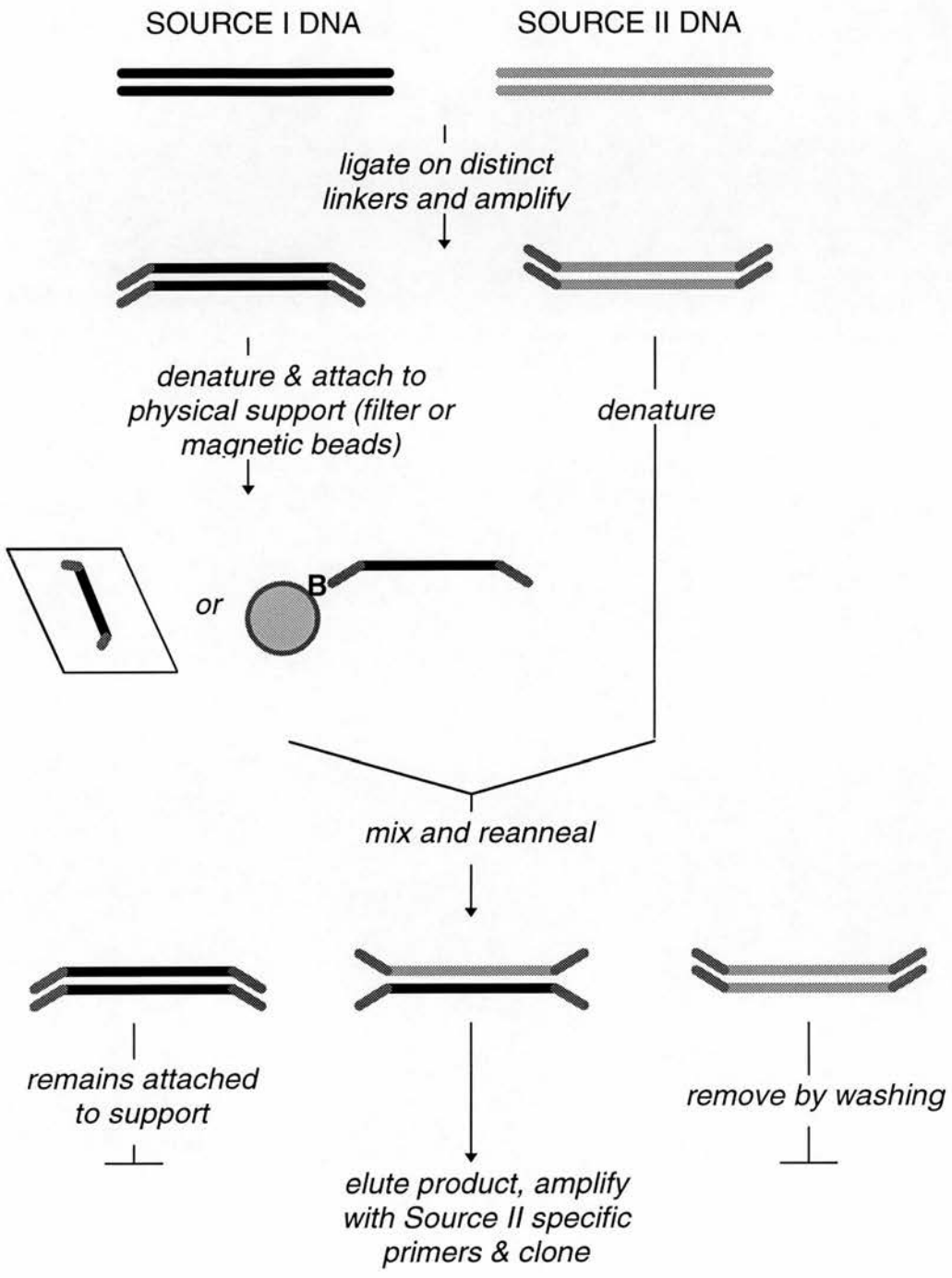


Fig. 5) Physical plus Ligation plus PCR selection ('End-Ligation Coincident Sequence Cloning')